REPORT

FINAL REPORT

Promoting Opportunity Demonstration: Design Report

(Deliverable 5.2)

May 18, 2018

David Wittenburg
Kenneth Fortson
David Stapleton
Noelle Denny-Brown
Rosalind Keith
David R. Mann
Heinrich Hock
Heather Gordon

Submitted to:
Social Security Administration
Office of Research, Demonstrations and Employment Support
4-P-8-D Annex
6401 Security Boulevard
Baltimore, MD 21235
Project Officer: John Travis Jones
Contract Number: SS00-17-60008

Submitted by:
Mathematica Policy Research
1100 1st Street, NE, 12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: David Wittenburg
Reference Number: 50390

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

CONTENTS

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

EXHIBITS

ACRONYMS

BOND	Benefit Offset National Demonstration
C	Control group
CFIR	Consolidated Framework for Implementation Research
IRS	Internal Revenue Service
MIS	Management information system
POD	Promoting Opportunity Demonstration
PROMISE	Promoting Readiness of Minors in SSI
RAPTER	Random Assignment Participant Tracking Enrollment and Reporting
SGA	Substantial Gainful Activity
SSA	Social Security Administration
SSDI	Social Security Disability Insurance
SSI	Supplemental Security Income
T1	Treatment group 1
T2	Treatment group 2
TWP	Trial Work Period
VR	Vocational Rehabilitation
WIPA	Workforce Incentive Planning and Assistance
YTD	Youth Transition Demonstration

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

EXECUTIVE SUMMARY

POD Evaluation Design Report (Deliverable 5.2) May 2018	
Background	• ***Benefit offset.*** Congress directed the Social Security Administration (SSA) to carry out the Promoting Opportunity Demonstration (POD) to test a benefit offset formula for Social Security Disability Insurance (SSDI) beneficiaries. POD tests a simplified set of work rules that replaces the cash cliff with a more gradual ramp, which we refer to as a benefit offset. The offset formula reduces benefits by $1 for every $2 of earnings in excess of a Trial Work Period level (defined as $850 in 2018). The evaluation of POD is a randomized controlled trial that tests two versions of POD rules and current rules for beneficiaries who volunteer for services. The two treatment arms use the same offset formula but differ in what happens to beneficiaries whose benefits are offset to zero. The evaluation also includes a control arm that is subject to current law. • ***Purpose.*** This report summarizes the evaluation design for POD. The evaluation will provide information on the outcomes for at least 9,000 subjects who meet the eligibility criteria, reside in select areas of eight POD states, and volunteer for the demonstration
Design of recruitment and implications for evaluation	• ***Enrolling volunteers and other demonstration features.*** The eligible beneficiaries who enroll are volunteers who self-select into the demonstration and do not represent the population of beneficiaries. In addition, POD includes provisions that allow treatment subjects to withdraw at any time and revert to current rules. Finally, the specific provisions of the POD offset rules could benefit some beneficiaries with certain characteristics, such as those who have completed the Trial Work Period, more than other beneficiaries. • ***Policy implications.*** The features of the demonstration represent important issues that our evaluation findings must address in interpreting findings for a national policy. For example, the volunteers will self-select into the demonstration, which has implications for generalizing estimates to a national policy.
Design of data collection and definition of outcomes	• ***Quantitative and qualitative data sources.*** Collecting data for the evaluation will include quantitative data (administrative data, management information system data, and surveys) and four rounds of qualitative data to support the analysis. • ***Primary outcomes.*** Across all analyses, we will emphasize four primary outcomes in our reports: earnings, benefit payments, substantive earnings, and income.
Design of analysis	• ***Process analysis.*** The evaluation's broad objectives include understanding the program environment; understanding the recruitment, enrollment, and random assignment processes; and describing and assessing implementation of the benefit offset and the associated services. • ***Participation analysis.*** The evaluation will calculate rates of enrollment, withdrawal from POD rules, use of the benefit offset and other services, and earnings reporting and overpayments, and document how rates vary with beneficiaries' characteristics. • ***Impact analysis.*** The design aims to provide rigorous estimated impacts of the benefit offset on substantive employment, earnings, SSDI benefits, and income from earnings and benefits combined, as well as impacts on other related outcomes to further explore the main findings. • ***Benefit-cost analysis.*** The evaluation will assess if the benefit offset and its associated components was cost beneficial when considering the combined benefits and costs to all beneficiaries and government entities.
Planned Reporting	• ***Four primary reports and eight special topic briefs.*** Our first two primary reports provide an early glimpse into recruitment, enrollment, and service activities through the first year of implementation. The next two primary reports will summarize program processes and outcomes through the end of the demonstration. We will develop eight policy briefs on special topics and issues that emerge during the study.

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

I. INTRODUCTION

As part of the Bipartisan Budget Act of 2015, Congress directed the Social Security Administration (SSA) to carry out the Promoting Opportunity Demonstration (POD) to test a benefit offset formula for Social Security Disability Insurance (SSDI) beneficiaries. The new rules attempt to simplify work incentives to promote employment, reduce dependency on benefits, and reduce administrative complexity. POD is part of a broader effort by policymakers to identify new approaches to help beneficiaries and their families increase their incomes and self-sufficiency through work. In addition to authorizing POD, the Bipartisan Budget Act extended the solvency of SSDI and renewed SSA's demonstration authority (United States Code 434). The renewed authority allows SSA to carry out experiments and demonstration projects that promote labor force attachment and identify mechanisms that could result in savings to the SSDI Trust Fund. Under the renewed demonstration authority, demonstrations must include participants who are volunteers and can withdraw at any time.

A potential challenge for beneficiaries to return to work is that existing work rules for SSDI beneficiaries are complex and include provisions that result in a loss of SSDI benefits for excess earnings (a phenomenon commonly called the "cash cliff"). One complexity is that current rules change over time depending on how the beneficiary's work profile changes. For example, the current rules do not result in any reductions for earnings among beneficiaries who initially return to work, which SSA refers to as the Trial Work Period (TWP). However, the rules change following the TWP (Exhibit I.1). Ultimately, SSDI beneficiaries who work over longer periods and earn wages above the Substantial Gainful Activity (SGA) threshold, defined in 2018 as $1,180 for non-blind beneficiaries and $1,970 for blind beneficiaries, risk the complete loss of benefits. Researchers and administrators refer to this benefit loss as a cash cliff because beneficiaries lose all benefits for a single dollar of earnings in excess of SGA.

The complexity of the current rules creates challenges for beneficiaries and SSA staff (Stapleton et al. 2006; Wittenburg et al. 2012; Hoffman et al. 2017). For beneficiaries, the complexity of the work rules creates fears about returning to work. Additionally, beneficiaries who do not fully understand the current rules risk incurring overpayments, which they then will need to pay back to SSA. Administratively, SSA staff must record beneficiary earnings, which can be complicated if beneficiaries do not regularly report their earnings to SSA.

POD attempts to address these challenges by creating a simplified set of new work rules that replaces the cash cliff with a "ramp," which we refer to as a benefit offset (see Exhibit I.1). Under POD, the rules do not change when the beneficiary's work profile changes. The new offset formula reduces benefits by $1 for every $2 of earnings in excess of a TWP level (defined as $850 in 2018), which we hereafter refer to as the POD threshold when referring to POD rules. We continue to refer to the TWP when referencing current rules.[1]

[1] As shown in Exhibit I.1, POD also includes special provisions for beneficiaries who have impairment expenses, which SSA refers to as Impairment-Related Work Expenses. SSA deducts approved Impairment-Related Work Expenses under current rules. We discuss the implications of the Impairment-Related Work Expenses in more detail in Chapter II.

Exhibit I.1. Snapshot comparison of current rules and the new POD rules

Current rules	POD rules
• SSDI beneficiaries can earn any amount during a TWP, which allows for earnings in excess of a threshold called the TWP threshold for nine months. • Additionally, the Extended Period of Eligibility, which lasts for 36 months after the end of the TWP, allows beneficiaries to remain on the program if they earn above the SGA. • During the Extended Period of Eligibility, there is a one-time exception to benefit reductions for earnings above SGA called the Grace Period. During this period, beneficiaries are eligible for benefits for the first month in which earnings exceed SGA and the following two consecutive months during the Extended Period of Eligibility. • However, following the Grace Period, beneficiaries face termination of benefits if their earnings exceed SGA.	• POD includes two treatment arms, both of which use the same new rules to calculate benefits. The new rules replace the cash cliff with a benefit offset, which reduces benefits by $1 for every $2 earned above the larger of the POD threshold (defined as $850 in 2018) and the amount of the POD subject's Impairment-Related Work Expenses. • The treatment arms differ in their rules governing termination of benefits for excess earnings. In both treatment arms, POD initially suspends benefits. However, in one arm the suspension is not time limited, whereas in the other arm, POD terminates benefits after 12 consecutive months of suspension.

POD tests two versions of the new work rules. Both versions of the treatment include the same offset rules noted above. Assuming a beneficiary has no Impairment-Related Work Expenses, under the new rules a beneficiary will never lose any benefits if earnings are less than $850 in 2018. For earnings greater than $850, a beneficiary will lose $1 of benefits for every $2 of earnings. The point of full offset under the POD rules depends on the beneficiary's benefit level. For example, beneficiaries who have monthly benefits of $1,100 will reach full offset under the new rules if they earn $3,050 or more (which equals the POD threshold of $850 plus two times the benefit level of $1,100). At this threshold, subjects in the first treatment group will remain eligible for benefits if their earnings decrease in the future. Subjects in the second treatment group will face termination of eligibility if their earnings remain at this level for 12 consecutive months. Similar to current rules, this second treatment group will be eligible for expedited reinstatement if subjects reapply for benefits after termination. The POD work rules include the new offset as well as other changes to current law rules, such as the elimination of the Extended Period of Eligibility and Grace Period and changes in rules for suspension and termination.

The evaluation of POD is a randomized controlled trial that tests these two versions of POD rules and current rules. Specifically, POD includes two treatment arms and one control arm for beneficiaries who volunteer to be randomly assigned to services (hereafter referred to as demonstration or study subjects). For shorthand, we refer to this study as an evaluation of the benefit offset, which encompasses the POD work rules, the POD benefits counseling, and the POD benefit offset itself.

SSA contracted with Mathematica Policy Research to lead the evaluation and Abt Associates to lead the implementation of POD in eight states over a five-year period (January 2017–December 2021). Mathematica will work with its partner, Insight Policy Research, to conduct an evaluation that will include comprehensive process, participation, impact, and cost-benefit analyses of all demonstration activities. Mathematica will also implement and manage

the study intake processes and collect data from multiple sources throughout the demonstration period.

For the implementation activities, Abt has developed an implementation team, which we refer to as the "implementation management and site staff," to deliver the associated services to support the implementation of new POD rules. The Abt implementation management and site staff includes Abt management to provide oversight for the implementation and several entities who will provide service delivery. The other entities include Vocational Rehabilitation (VR) agencies in four states (Alabama, Connecticut, Maryland, and Vermont) and Work Incentives Planning and Assistance agencies in four other states (California, Michigan, Nebraska, and Texas), who will coordinate the delivery of POD-related supports and support SSA in gathering the needed information from demonstration subjects in administering the new offset rules. These agencies will fill the POD work incentive counselor positions internally or will contract with local vendors (such as community rehabilitation programs) to provide the counseling services. Abt has established a POD call center in McAllen, TX, to respond to calls from treatment subjects, SSA, project partners, and the general public. Virginia Commonwealth University will provide technical support to the implementation team. Finally, Abt will have three support units (POD Central Operations, POD Processing Center, and POD Earnings Support) responsible for providing indirect support to treatment subjects and for supporting the work incentives counselors in all eight sites.

This report summarizes the evaluation design for POD. The primary objective of the report is to serve as a reference for government officials, demonstration staff, providers, and members of the general public to consult for basic information about POD until reports become available later in the evaluation.

A. Lessons from recent offset tests: Benefit Offset National Demonstration

The POD evaluation will draw on lessons from previous related work incentive experiences, most notably SSA's Benefit Offset National Demonstration (BOND), which is testing a different offset formula. Similar to POD, BOND included a $1-for-$2 offset formula. However, in contrast to POD, the BOND formula (1) included provisions for the TWP and Grace Period and (2) had a higher earnings threshold before starting the offset based on an annual total (defined as 12 times the SGA level). Early results suggest the BOND offset has had limited impacts on earnings in the short term, and benefit payments have increased in the short term (Wittenburg et al. 2015; Gubits et al. 2014). The evaluation findings also found administrative delays in making benefit adjustments following changes in earnings. These latter findings underscore some of the complexities of the BOND rules. Specifically, the BOND rules keep elements of current law (for example, the Trial Work Period and Grace Period) and include an annualized version of earnings (annual SGA) that potentially added to the delays given the complexity of these calculations.

POD addresses the perception that BOND rules were complex by using a simplified set of administrative adjustments in implementing the offset. Specifically, the POD rules eliminate the TWP and Grace Period, which effectively make the POD offset rules consistent throughout the entire demonstration period for all beneficiaries unless their benefits are terminated. This change differs from current rules (and the rules under BOND) because beneficiaries experience different

work incentive rules following the Grace Period. To expedite the processing of the POD rules, the state VR and Work Incentive Planning and Assistance agencies noted above will collect and coordinate the subjects' earnings information and transmit the information to SSA for timely benefit adjustments. These agencies will also likely be the primary source for benefits counseling for the POD rules and employment services. Finally, all else equal, because the POD offset uses a threshold for calculating the offset in comparison to BOND (that is, a monthly POD threshold in comparison to an annualized SGA threshold that only applies after the Grace Period), we expect there will be more offset users in POD.

B. POD theory of change and evaluation research questions

The "theory of change" for POD is shown in Exhibit I.2, which summarizes the intervention components for two treatment groups and the potential beneficiary and system outcomes from successful implementation of demonstration services. In the near term, the POD interventions could increase work-related activity, such as counseling, job search, use of other employment services, and efforts to increase earnings at existing jobs. Over the long term, participants could experience increased employment and earnings, benefit reductions, and increases in total income due to increases in earnings that more than offset benefit reductions. Those changes in earnings, employment, and income could, in turn, lead to improved health statuses and quality of life.

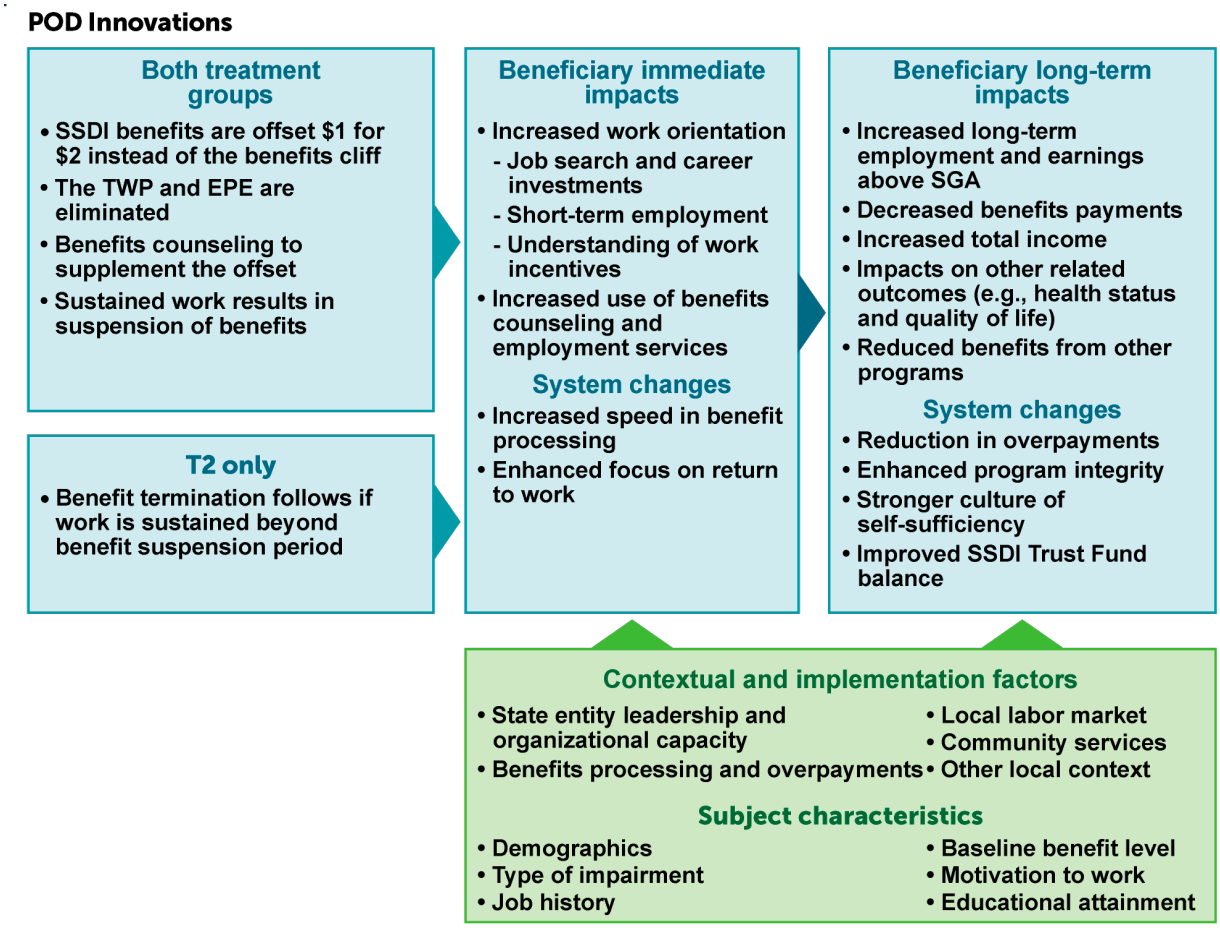
The POD interventions could also influence system changes that improve SSA's administrative processes, including reducing the time it takes to apply the offset to benefits. These changes could reduce the occurrence of overpayments, reinforce participants' understanding of how the offset works, increase beneficiaries' trust, and otherwise enhance program integrity. If the demonstration leads to a reduced reliance on benefits and improvements in system outcomes, it could also generate savings for the SSDI Trust Fund as well as savings for other social safety net programs.

Building from the theory of change, we designed the POD evaluation to answer the following high-level research questions:

- What are the impacts of each of the two sets of POD rules on beneficiaries' earnings, SSDI benefits, and total earnings and benefit income? We will use the impact analysis (Chapter VIII) to answer these questions.
- Is POD attractive to beneficiaries, particularly those whose earnings and benefits would most likely be affected under the POD $1-for-$2 offset? Do they remain engaged over time? We will use the process (Chapter VI) and participation (Chapter VII) analyses to answer these questions.
- How were the POD offset policies implemented, and what operational, systems, or contextual factors facilitated or posed challenges to administering the offset? We will answer these questions through the process analysis (Chapter VI).
- How successful were POD and SSA in making timely benefit adjustments, and what factors affected timeliness, positively or negatively? We will also address these questions as part of the process analysis (Chapter VI).

- How do the impacts of the benefit offset policies vary with beneficiary characteristics? Answering this question will be part of the impact analysis (Chapter VIII).
- What are the benefits and costs of the benefit offset relative to current law, and what are the implications for the SSDI Trust Fund? This is the focus of the benefit-cost analysis (Chapter IX).
- What are the implications of the POD findings for national policy proposals that would include a SSDI benefit offset? Answering this question will draw mainly on findings from the impact analysis supplemented with findings from the process and participation analyses.

Exhibit I.2. Theory of change for the POD evaluation



EPE = Extended Period of Eligibility

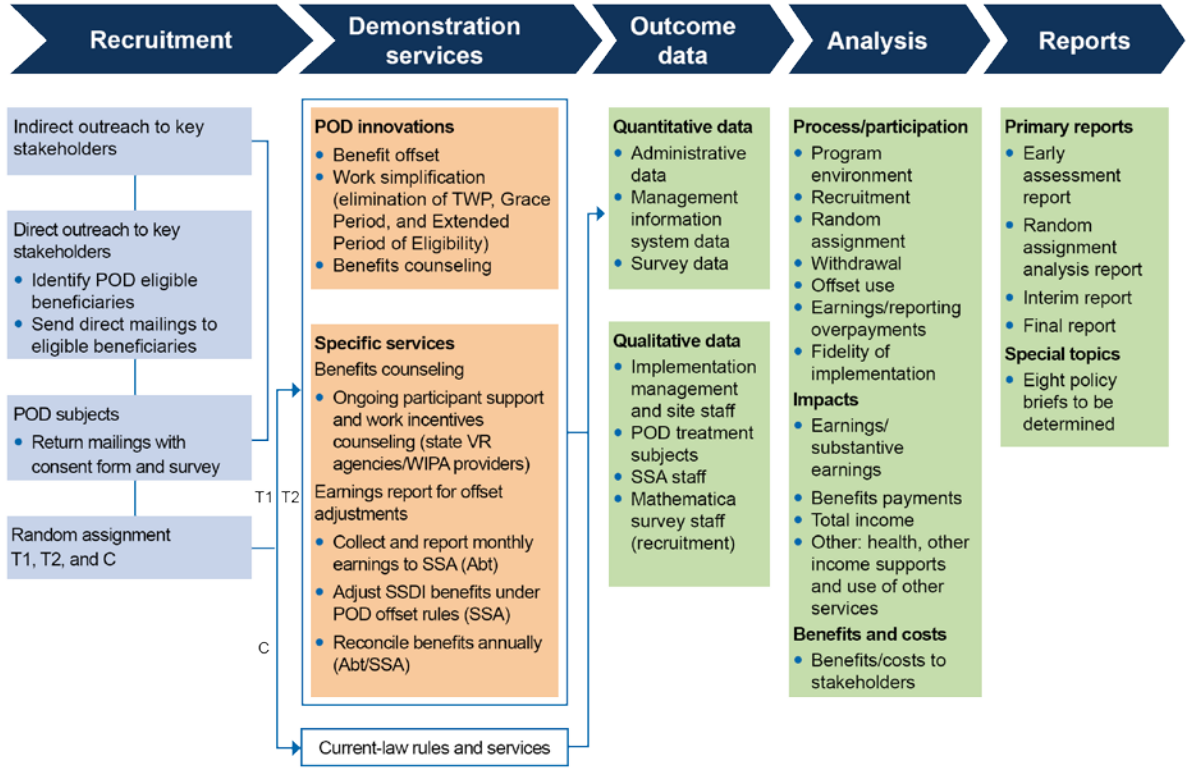
T2 = Treatment group 2. Treatment group 2 has the same offset rules as treatment group 1 except they are subject to benefit termination for excess earnings.

C. POD Evaluation framework

We conclude with a summary framework for the evaluation, which we use to motivate the structure of this report (Exhibit I.3). The blue arrows across the top of the figure display five key elements of the framework (recruitment, demonstration services, outcome data, primary analysis and special topic analysis). The colors of the boxes under the blue arrows summarize the

different components of the evaluation. The blue boxes (recruitment) summarize interactions between the evaluation team and beneficiaries during the recruitment phase. The orange boxes (demonstration services for treatment subjects) describe the rules and services applicable to treatment subjects. There is also an unshaded box under demonstration services, which represents the current-law rules and services for control subjects. Finally, the green boxes (outcome data, primary analysis, and special topic analysis) detail features of the evaluation itself.

Exhibit I.3. Evaluation framework



Demonstration services. In Chapters II and III, we describe the demonstration services and the theoretical implications of POD for potential outcomes in the evaluation. In Chapter II, we focus on predictions that draw on empirical evidence from BOND and economic theory, and how we will interpret the impact findings. In Chapter III we summarize the rules and services that will apply to treatment subjects during the demonstration, including how the rules differ for T1 and T2 subjects with respect to termination for work, and the benefit offset and the work provision rules.[2] We also provide an overview of the specific services to support implementation, including benefits counseling and earnings reporting.

Recruitment. In Chapter III, we describe our general approach to recruitment. This approach will first involve conducting indirect outreach to key stakeholders to inform them about POD so that they too can serve as a resource for beneficiaries with questions. We will then conduct direct

[2] More detail on demonstration services appears in the POD Implementation Design Report (Abt Associates 2017a).

outreach by sending mailings to eligible beneficiaries. Those who volunteer to participate must sign a consent form and complete a survey. Finally, we will randomize eligible subjects to one of three study groups.

Outcome data. In Chapters IV and V, we describe the quantitative and qualitative data we will include in the analysis, respectively. In Chapter IV, we summarize the three components of quantitative data that we will use in our analyses (administrative data, management information system data and surveys). In Chapter V, we describe the sources of qualitative information (implementation management and site staff, POD treatment subjects, SSA staff, and Mathematica survey staff) that we will gather during four rounds of data collection. Taken together, we will use data summarized in these chapters to inform the analyses in the subsequent chapters.

Analysis. In Chapters VI, VII, VIII, and IX, we describe our methodological approach. In Chapter VI, we summarize our approach to the process analysis, which includes a description of the implementation process and assessment of implementation fidelity to the design of the demonstration. This analysis will provide formative feedback to refine POD implementation and inform the interpretation of evaluation findings. In Chapter VII, we describe the participation analysis, which will summarize the characteristics of SSDI beneficiaries who do and do not volunteer for POD, as well as differences in participant characteristics and participation in POD by treatment arms. In Chapter VIII, we describe the impact analysis, which will be our approach to estimating impacts among POD participants in each POD treatment arm and the control group, as well as assessing impacts among subgroups. In Chapter IX, we describe the benefit-cost analysis, which will assess the costs and net benefits of the POD demonstration on POD beneficiaries, SSA, other government agencies, and society.

Reports. In Chapter X, we summarize the content and timelines for the four primary reports and eight special topic briefs. Our first two primary reports (an early assessment report and recruitment and random assignment analysis report) will provide an early glimpse into recruitment, enrollment, and service activities through the first year of implementation. Our second two primary reports will summarize program processes and outcomes at the midpoint and end of the demonstration. We will develop eight policy briefs that extend the information in these reports to address emerging topics of interest during the demonstration.

Addendum: The addendum includes a summary of updates for the recruitment described in Chapter III that were implemented following the pilot period. During the pilot period, SSA and the evaluation team worked to refine the recruitment processes, which included adding additional counties to one of the eight states (Texas) and changes in other recruitment processes. Readers interested in the revisions to the recruitment process outlined in Chapter III, including the final set of areas included in POD and outreach methods to recruit beneficiaries, should refer to the addendum.

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

II. CONCEPTUAL FOUNDATION FOR POD

This chapter reviews recent findings and a summary of economic theory related to implementing an offset for beneficiaries that are relevant to understanding the logic of POD and potential outcomes for the evaluation. We first review findings from the evaluation of BOND, which provides some context for the theoretical discussion of POD and potential outcomes given that both BOND and POD included an offset design. Next, we present four primary outcomes for POD and discuss predictions for these outcomes from economic theory. Appendix A includes more detailed information about the models we used to develop predictions. Finally, we review the specific implications for potential outcomes in the evaluation.

A. Findings from BOND

As noted in Chapter I, the POD design built on lessons learned from the SSA's BOND evaluation. Similar to the POD benefit offset, the BOND benefit offset reduced monthly benefits by $1 for every $2 of earnings above a threshold. The BOND formula differed from POD in that BOND included provisions for the TWP and Grace Period and had a higher earnings threshold before starting the offset based on an annual total (defined as 12 times the SGA level).

We provide more detail on the BOND findings available to date and how those findings have influenced the design of the POD offset. The impact findings also provide lessons and useful benchmarks for the evaluation of POD.[3]

The BOND evaluation features two stages. For Stage 1, SSA sought to provide direct estimates of the impacts of the BOND offset on the national beneficiary population. SSA designed Stage 2 of BOND to examine impacts among those beneficiaries who seemed most likely to use the offset, informed SSDI-only volunteers, and to test significant enhancements to the counseling provided to the Stage 1 treatment subjects. A research team from Abt and Mathematica is conducting the evaluation, which is approaching its final stages with a draft of the final report due in spring 2018. The findings we cite include information from the most recently released reports: Hoffman et al. (2017) and Croake et al. (2017) for Stage 1 and Geyer et al. (2017) for Stage 2.

Participation in BOND Stage 1 was mandatory, and the Stage 1 sample included a nationally representative sample of the SSDI beneficiary population who were over age 20 by September

[3] SSA also previously conducted the Benefit Offset Pilot Demonstration as a pilot test for BOND, which used the same benefit offset formula but with some differences in the earnings to which the offset applied and administrative details. Weathers and Hemmeter (2011) reported mixed findings of the impacts of the offset on earnings and benefit amounts. They found a significant increase in the percentage of treatment group subjects earning above annualized SGA. However, they found virtually no difference in mean earnings one year after random assignment and a modest difference in the second year (slightly less than $1,000 per year) that was not statistically significant. They also found a significant increase in mean benefits of about $500 in each of the first two years after random assignment. The authors pointed to shortcomings and delays with the processes used to report earnings, complete work Continuing Disability Reviews, adjust benefits, and reconcile benefits at the end of the year. The BOND evaluation built on these lessons, which is why we focus on BOND findings here. For a more detailed review of the Benefit Offset Pilot Demonstration, as well as other more general incentive initiatives such as the Negative Income Tax experiment, see Bell et al. (2011) and Delin et al. (2010).

2017 and younger than 60 at the start of BOND. Specifically, the research team randomly selected the sample from 10 SSA Area Offices around the country that were themselves randomly selected from an exhaustive set of 53 SSA Area Offices. The research team used program records to randomly select a treatment group (almost 80,000 subjects) and a control group of more than 900,000 from all SSDI beneficiaries in the 10 areas.

The study sample for Stage 2 comprises informed SSDI volunteers, similar to POD, except that BOND Stage 2 includes SSDI-only beneficiaries and excludes those who are concurrent beneficiaries. The counseling enhancements provided to Stage 2 treatment group members go substantially beyond the standard counseling services provided under current law (including to the BOND control groups) and to the Stage 1 treatment subjects, featuring proactive initial and follow-up outreach by the counselors and several enhancements to the counseling itself. According to Gubits et al. (2014), the volunteer rate was 5.4 percent in Stage 2.[4] The study sample in Stage 2, similar to POD, represents a volunteer sample.

The BOND research team specified confirmatory outcomes that included average SSDI benefits paid and annual earnings, and exploratory outcomes that included the remainder of outcomes (see Bell et al. 2011). They selected the benefits paid and earnings outcomes they considered most important for both theoretical and policy reasons. The confirmatory outcomes received the most prominence in the reports and the research team used specific statistical adjustments to present the findings.[5]

Through 2015, the fifth year of the demonstration, for the confirmatory outcomes, the BOND Stage 1 evaluation has found no evidence of an impact on the average annual earnings and positive impacts on average SSDI benefits. The exact benefit estimate will change when retroactive changes are made for 2015 benefits, but it is unclear whether the estimate will increase or decrease.

The Stage 1 evaluation shows positive impacts on some notable exploratory outcomes with substantive earnings; it found an increase in annual earnings above the annualized SGA amount (an increase of 9.1 percent relative to the control group), but it also found evidence of reductions in the percentages with earnings above twice that amount and above three times that amount (by 9.0 percent and 11.3 percent relative to the control group, respectively). The evidence of negative impacts for the percentages earning above two or three times the annualized SGA indicates that the beneficiaries who would have had their benefits suspended or terminated for work under current law experienced windfall gains. Those beneficiaries, on average, earned less than they would have earned under current law, offsetting increases in mean earnings for those induced to

[4] We developed this estimate by calculating the total number of volunteers enrolled (12,954) divided by solicited beneficiaries (238,070).

[5] Specifically, the research team presented statistically significant findings from confirmatory analyses "as the proven impacts of BOND without a multiple comparisons problem. In contrast, statistically significant findings from exploratory analyses that do not adjust for multiple comparisons must be characterized as simply suggestive of what BOND can accomplish, but not proven" (Bell et al. 2012, page 132). In Chapter IX, we discuss our approach to addressing the multiple comparisons issue based on lessons from BOND and recent developments from the literature, designating outcomes as "primary" or "secondary," which are related to the concepts of confirmatory and exploratory outcomes but do not imply multiple comparisons adjustments.

earn more, and more than offset reductions in the latter group's benefits. These exploratory findings are of note because they align with theoretical predictions of how beneficiaries would respond to the offset. In addition, the findings provide context to the variation in beneficiary earnings responses across different earnings levels, which is a notable issue that we discuss in the following POD theory section.

Earnings and benefit impacts for the BOND Stage 2 treatment subjects potentially differ because these subjects are informed SSDI-only volunteers, but to date they are qualitatively similar to those for Stage 1. For the treatment group that, like the Stage 1 treatment group, received BOND's standard counseling services, the increase in mean earnings in 2015 was not statistically significant. The impact on benefits paid in 2015 was significant, positive, and larger as a percentage of control group benefits paid than for the Stage 1 estimate: 4.3 percent. As with Stage 1, the research team expects the benefit impacts to change somewhat as SSA retroactively adjusts the benefits. The researchers found evidence of an increase in the percentage with earnings above the annualized SGA amount that was larger relative to the control group percentage than the increase for Stage 1: 26.4 percent. Unlike with Stage 1, there was no statistically significant evidence of reductions in the percentages with earning above two or three times the annualized SGA amount. Thus far, the evaluation has found that impacts for those assigned to the enhanced-counseling treatment group do not differ significantly from those for the standard counseling treatment group.

The BOND evaluation also included some important survey findings regarding beneficiaries' knowledge about the new rules and offset adjustments that are relevant to the POD evaluation. Of particular note is the finding that the understanding of the offset among Stage 2 treatment subjects was higher than among Stage 1 treatment subjects. Twenty-nine percent of Stage 1 treatment subjects gave survey responses consistent with a correct understanding of how earnings affect benefits under the BOND offset. This is compared to the 48 percent of Stage 2 treatment subjects—recruited and informed volunteers—who were offered the same benefits counseling available to Stage 1 treatment subjects. However, both groups of treatment subjects' understanding was less than the 54 percent of control subjects (in both stages) who exhibited a correct understanding of how their earnings affect their (current law) benefits. This suggests that the recruitment and composition of the sample as well as the complexity of the rules governing earnings may affect beneficiaries' understanding. Hence, this suggests tracking beneficiaries' understanding of the POD rules will be an important component of the POD evaluation.

The BOND findings about the significant delays in the benefit adjustment process is also notable, given it might have created some confusion for beneficiaries about the BOND rules. Further, many treatment subjects, especially in Stage 1, did not take advantage of the special process for reporting earnings and initiating benefit adjustments. Resource constraints at SSA resulted in long delays in the processing of work continuing disability reviews, even among those who reported their earnings, and significant functional limitations of a software program developed to automate a large share of adjustments resulted in more delays, many manual adjustments, and diverting resources from the processing of work continuing disability reviews.

In summary, although there are important differences between the BOND and POD offset rules, the lessons from the BOND evaluation have important implications for components of the POD evaluation. The specification of confirmatory outcomes is relevant to the theory of POD,

whereas the early impact findings might provide some clues to issues our team should emphasize in the evaluation.

B. Overview of POD primary outcomes and theory

The theory of change for POD outlined in the introduction includes potential impacts on several outcomes, though for reporting purposes, it is important to identify primary and secondary outcomes for the evaluation. As explained more thoroughly in Chapter VIII, we differentiate between the primary and secondary outcomes to distinguish the measures that should receive the most policy focus in the ultimate evaluation of the benefit offset's efficacy. This designation is a transparent way to avoid concerns about data mining when assessing impacts on the broad range of outcomes. Next, we summarize primary outcomes for the evaluation and discuss the theoretical effects of the POD offset and associated rules on these outcomes.

We recommend four primary outcomes, including the two confirmatory outcomes from the BOND evaluation (earnings and benefits) noted previously and two additional outcomes (substantive earnings, defined as earnings above SGA, and income, defined as earnings plus benefits) for the POD evaluation. The two we draw from the BOND evaluation are policy relevant in their own right and facilitate comparisons across studies. We will include substantive earnings and income as primary outcomes in POD because, as we will discuss, both are tied to the theoretical considerations for impacts and thus are strong indicators of program success. We can measure all four primary outcomes using administrative data (Chapter IV), which is important to tracking outcomes beyond the demonstration.

Theoretical considerations related to primary outcomes. An important starting point for considering theoretical outcomes for POD is to define income and substitution effects. The previous economic literature has shown that a change in program, tax, or other rules that govern the relationship between an individual's earnings and income, such as a benefit offset, has an income and a substitution effect. The POD rules have countervailing income and substitution effects similar to a reduction in a wage tax, except that the predictions also depend on the exact earnings and benefits levels of the study subject. The income effect is negative: if benefits increase and thus income increases, the theoretical expectation is that earnings will decline because of a decline in the value to the individual of the last dollar earned. One recent study produced an estimate of this effect for SSDI beneficiaries: a $1 increase in income from SSDI benefits reduced an average beneficiary's earnings by $0.20 (Gelber et al. 2017). The substitution effect is positive if an individual will convert additional dollars of earnings to income at a higher rate because of the change. Depending on the nature of the rules change, these two effects either reinforce each other (yielding a prediction about the direction of earnings change), or compete with each other (yielding a prediction that is ambiguous). We provide a summary example describing the economic theory for income and substitution effects in the text box for reference in the following discussion. (See Borjas 2016 for a detailed summary of income and substitution effects.) In Appendix A, we cover the theory in more depth to support this overview.

Example: The income and substitution effects of a reduction in a wage tax

Holding earnings constant, a reduction in a tax on an individual's wages increases both the individual's income and the rate at which an increase or decrease in earnings is converted to a change in income. The **income effect** is the result of the first change. As income increases, the incentive to work decreases because the value to the individual of the last dollar of earnings relative to the value of the last hour spent working is presumably lower than what it was before. Hence, the income effect alone leads to an expectation of reduced earnings—a negative effect. The **substitution effect** is a result of the second change, and it is positive. The reduced wage tax example increases the incentive to work at the rate at which earnings are converted to income increases. Hence, the substitution effect alone predicts an increase in earnings. In this case, the two effects work in opposite directions, and the predicted direction of the change in earnings is ambiguous.

In POD, the predicted effects for the four primary outcomes are ambiguous in three cases (earnings, benefits, and income) because of the counteracting forces of the income and substitution effects. The predicted effect of the POD rules on mean earnings is ambiguous because, although the POD offset creates attractive opportunities to earn more for some beneficiaries (positive substitution effect), other beneficiaries might have incentive to earn less because of countervailing income effects. Examples of beneficiary groups who might earn less include those who would work and not give up their benefits under current law (such as those who earn between TWP and SGA) and those who find it more attractive to reduce their earnings and increase their benefits under the POD offset (such as those with very high earnings well above SGA who would receive benefits from the POD offset ramp). For benefit payment outcomes, the ambiguity in the aforementioned earnings effects creates ambiguity in predicted benefit payment outcomes given that earnings and benefits are directly related through the offset. Similarly, there is an ambiguity in income given the combination of ambiguous effects for benefit payments and earnings. In Appendix A, we review the economic theory associated with the POD offset for those interested in more detail on these predictions.

Summary of theoretical predictions (detailed in Appendix A)

- **Mean earnings (ambiguous) and annual earnings above the annualized SGA amount (positive).** As under the BOND offset, some beneficiaries will earn more because the POD offset creates attractive opportunities to earn more, but increases in their earnings might not be large enough to overcome decreases in earnings for the substantial number who will find work effort less attractive under the POD offset.
- **Mean benefit payments (ambiguous).** Beneficiaries could receive benefits in POD, while under current rules they would receive zero benefits. However, the benefit increase might not be enough to offset the lower benefits paid during the TWP and Grace Period months that do not exist in POD and, in later months, for earnings between the TWP amount and the sum of the SGA amount plus any Impairment-Related Work Expenses they might have under current law.
- **Mean beneficiary incomes (ambiguous).** The predictions for beneficiary income are directly tied to the benefit payments and earnings predictions above.

These predictions ignore several factors that might affect impacts, such as the simplification of administrative rules and changes in termination.

The one outcome for which there is a positive prediction for earnings is for the substantive earnings outcome (defined as earnings above SGA). Specifically, there is a positive association with the POD rules and substantive earnings above SGA, given the offset removes the cash cliff. However, for substantive earnings well beyond SGA, an income effect could dampen the effect on substantive earnings.

Dynamic implications of changes in beneficiary perceptions about reduced uncertainty with benefit adjustments. An important issue not accounted for above is how simplifying the POD rules might reduce uncertainty for beneficiaries and potentially increase their willingness to work. Under current law, benefit adjustment depends on whether a beneficiary has completed his or her Grace Period. In POD, the benefit offset is the same regardless of timing. Hence, POD could increase beneficiaries' willingness to work more by reducing the uncertainty of benefit suspension or termination associated with current law. This factor would likely be strongest for those who currently earn below SGA or are not working, suggesting this factor could increase employment and earnings outcomes. The overall effects on all outcomes is still ambiguous because we do not have information on the importance of this factor relative to the other factors above. In the participation analysis, we will assess beneficiary perceptions of the offset to gain a better understanding of how well they understand the new rules and whether this understanding changes their perceptions about work.

A related feature of POD for the first treatment group is the elimination of the termination provisions. This feature of POD could further reduce the uncertainty that beneficiaries face in making work decisions. For example, if POD changes beneficiary perceptions about loss of benefits, POD could lead to employment increases beyond those described above.

C. Interpreting evaluation findings

A major goal of the POD evaluation is to make inferences about those impacts based on the demonstration. Multiple features of the demonstration itself will lead to substantial differences between impacts for demonstration treatment subjects and impacts that would occur under a national program. Most notably, the demonstration will use volunteers from a random sample of all SSDI beneficiaries, requiring them to complete an informed consent process before volunteering. In addition, those volunteers assigned to one of the two treatment groups have the option of reverting to current law, and community knowledge about how earnings affect SSDI benefits will be grounded in current law rather than the new rules. Demonstration processes for earnings reporting and benefit adjustments might differ from those in a national program. In this section we discuss in more depth how these features complicate extrapolating findings from the conceptual framework underpinning POD to a national policy. The participation analysis, detailed in Chapter VII, will be crucial for helping draw inferences about national policy from the demonstration estimates.

Implications of including volunteers. Our expectation is that informed beneficiaries will volunteer only if they believe that assignment to one of the treatment groups will benefit them—that over the course of the demonstration they will work and earn enough to be better off under the POD design than they would be under current law. Volunteers' expectations about future earnings could be wrong, and their understanding about the POD design less than complete, so at least some volunteers will ultimately not benefit from the POD design. Nonetheless, it is likely that eventual POD subjects will disproportionately include those who are most likely to benefit from the new rules, including those with relatively higher benefit amounts, more substantial earnings at baseline, no SSI or other benefits (such as private disability insurance), and who are

near or past their Grace Period. These individuals will have stronger incentives to volunteer for POD relative to other beneficiary groups.[6]

Beneficiary interest in volunteering will vary by subgroup

Volunteer rates should be relatively higher among beneficiaries with the following characteristics at baseline:

- **High benefit amounts.** Beneficiaries with high benefit amounts will have a longer benefit offset range and relatively more to lose from the current law cash cliff than beneficiaries with more limited benefits. In addition, the relatively high benefit amounts signal high pre-SSDI earnings, making it more likely that the beneficiary will have relatively more employment opportunities than other beneficiaries.
- **More substantive employment and earnings.** Beneficiaries who are already working will volunteer at higher rates than those who are not working because they are more likely to take advantage of the POD offset quickly.
- **No SSI benefits or other benefits (such as private disability benefits).** Work rules for SSI and other programs reduce any gains available under the POD design. At the extreme, more earnings or income could negate all other benefits and loss of other associated supports (for example, Medicaid).
- **Near or past their Grace Period.** Beneficiaries who are beyond the Grace Period could benefit from staving off benefit termination if they continue to engage in SGA. The POD rules might be attractive to beneficiaries who have more TWP and Grace Period months behind them because they potentially would lose less under the POD design after the Grace Period in comparison to current rules.

The differences in volunteer rates by characteristics and, specifically, our inability to observe impacts for nonvolunteers, is important for interpreting impact findings, especially in generalizing findings to a national policy. For instance, theory predicts that some beneficiaries who earn between the TWP (POD threshold) and the SGA amounts under current law will choose to work less under POD, but we might not observe such behavior among volunteers, because such beneficiaries are unlikely to volunteer in the first place. Hence, projections of impacts for a national program must adjust for potential unobserved impacts for this group, which is challenging because the incentives presented to nonvolunteers might theoretically differ from the incentives available to those who volunteer.

Implications of voluntary withdrawal options for treatment subjects. Under the demonstration, POD treatment subjects can withdraw from the demonstration and revert to current law at will. This protects those volunteers who find that, despite expectations, they are worse off under the POD offset than under current law, and presumably will increase the volunteer rate relative to what it would be without this provision. All else equal, we expect withdrawals from the demonstration will be higher for treatment subjects who (1) have more TWP and Grace Period months remaining under current law and (2) for T2 subjects relative to T1 subjects because the termination provisions exist in T2 (as they do under current law). The incentives for T2 subjects to withdraw are especially important because it is possible some might enter the demonstration with the opportunity to receive the protection from termination under the T1 treatment and subsequently drop out if they are randomly assigned to T2. In general, an increase in withdrawal will dampen impact estimates because those who revert will no longer

[6] In addition, we expect that beneficiaries who have a severe vision impairment and, hence, a higher blind SGA amount will have less incentive to participate given the advantages of the higher SGA amount under current rules relative to the POD threshold for the TWP amount.

receive the treatment. Although theory predicts some groups might have relatively higher dropout rates, no empirical evidence indicates how much of a problem withdrawal rates will create for the evaluation.

Duration differences in the availability of new POD rules. The demonstration offers treatments subjects assigned to a treatment group the opportunity to use the POD design for a limited period only, though the duration depends on when a subject enrolls in POD. Specifically, as outlined in the next chapter, POD recruitment will occur over a 12-month period. All else equal, beneficiaries in the earlier part of the recruitment period will have more exposure to the new rules than those at the end of the period.

The time-limited duration and the rolling nature of recruitment with a fixed end date on service delivery have important implications for who volunteers and ultimately for the interpretation of findings for a national policy. Specifically, some beneficiaries who would benefit from the POD design if available permanently will not volunteer, because they will not be able to benefit during the demonstration period. The most apparent reason is that they might have TWP and Grace Period months left, in which case the POD design will reduce their income in the months remaining. Hence, it would only make sense to volunteer if they thought they could more than recoup those losses in the post-Grace Period; the shorter the post-Grace Period, the less time to recoup the losses. In addition, beneficiaries who are not working or who have part-time work might need time for training and finding a new or better job to take advantage of the POD design. Hence, beneficiaries who might make such investments under a national program might not volunteer for POD.[7]

Summary. Including volunteers, options to withdraw, and variations in the duration of exposure to benefits are important issues that our evaluation findings must address in interpreting findings. We will refer back to these issues, particularly in the process and participation analysis (Chapters VI and VII) to document who volunteers, who withdraws, and the dosage of services received. We will then use this information to make cross-cutting assessments to inform the interpretation of impacts (Chapter VIII), particularly in generalizing findings for a national POD policy.

[7] We can test the following prediction by analyzing how particular baseline factors affect the likelihood of participation, which is part of the participation analysis discussed in Chapter VII: *The longer the remaining POD participation period at the time of recruitment, the more likely beneficiaries will volunteer.* The reason is the effect on the number of months under which the enrollee can potentially benefit from the POD offset. A positive finding from such a test would imply that a long-term, permanent POD offset would be more valuable to beneficiaries than one that is available for only two to three years.

III. TREATMENT AND CONTROL CONDITIONS, STUDY ENROLLMENT, AND RANDOM ASSIGNMENT

In this chapter, we describe the approach to identifying and randomly assigning beneficiaries into the following three study groups:

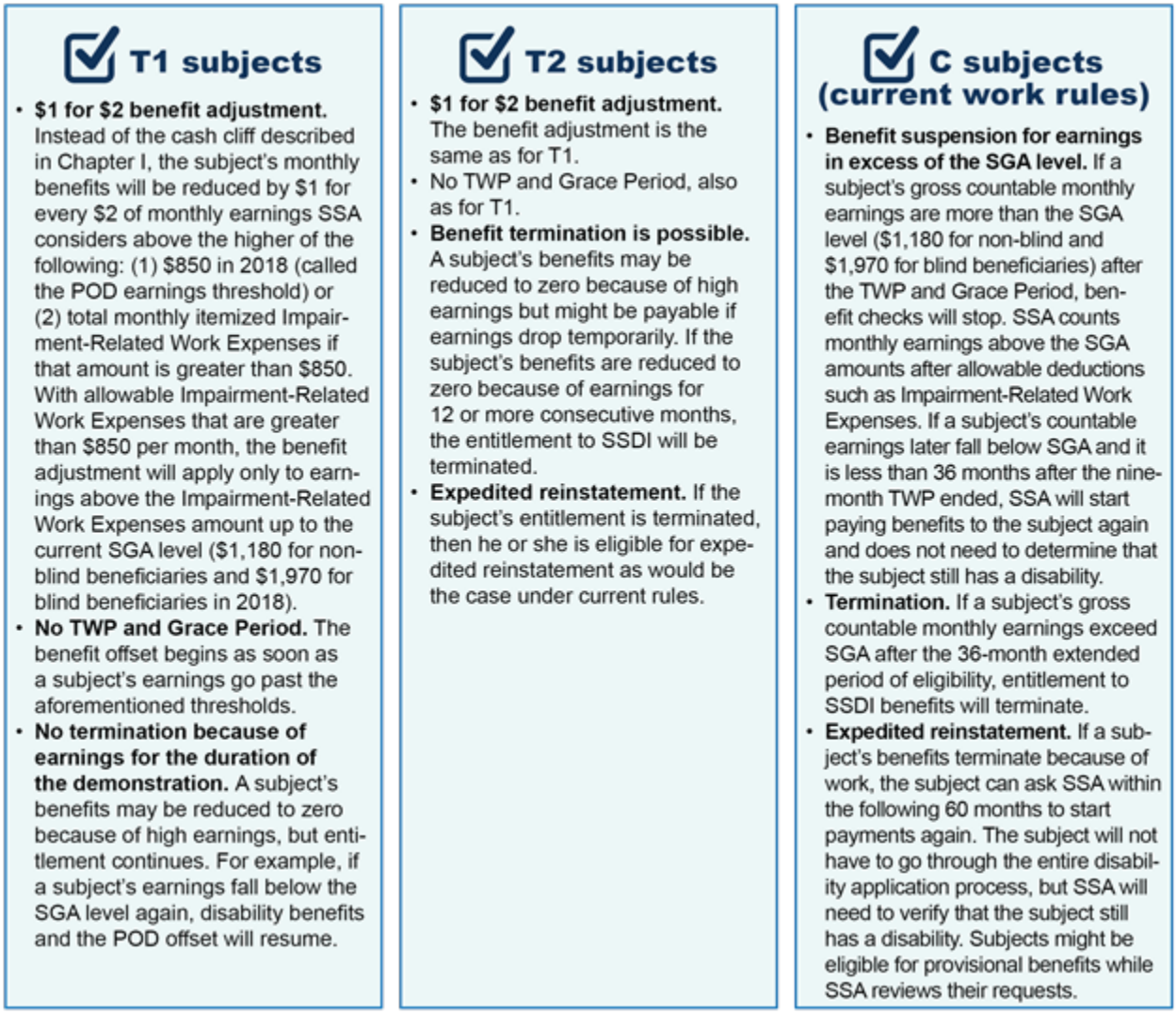
- ***Treatment group 1 (T1) subjects:*** T1 subjects will have a benefit offset of $1 for every $2 earned above the larger of the Trial Work Period level and the amount of the beneficiary's Impairment-Related Work Expenses and will not face benefit termination if their earnings reach the full-offset level (that is, go to zero). For a beneficiary with a monthly benefit of $1,100, the full-offset level in 2018 would be $3,050. If their earnings reach the full-offset level their benefits will be suspended, but benefits would resume if their earnings decreased below the full-offset threshold before the end of the project.
- ***Treatment group (T2) subjects:*** T2 subjects have the same benefit offset rules except, they do have eligibility provisions that result in benefit termination if they have earnings above the full-offset threshold for 12 consecutive months.
- ***Control (C) subjects:*** Control subjects will operate under current work rules, which include the "cash cliff," TWP, extended period of eligibility, and benefit termination.

We first summarize how the POD rules compare to current work rules, and the implications for our study sample. We then describe our recruitment process to identify eligible beneficiaries and randomly assign them to the three study groups above. We then describe our process for monitoring recruitment and random assignment, which will be essential to meeting SSA's sample targets for the demonstration. The issue of meeting sample targets is especially important given we do not know how many beneficiaries might be interested in volunteering in the demonstration. Hence, our approach to monitoring recruitment includes strategies for identifying and encouraging more beneficiaries to volunteer to participate, especially early on in the demonstration, using a pilot period to assess initial volunteer rates and, as necessary, make adjustments. Finally, in later sections, we conclude with a summary of our recruitment approach for the interpretation of evaluation findings, which is especially important in generalizing the findings of POD to a broader SSDI beneficiary population. In the addendum, we provide updates to this chapter based on lessons learned during recruitment.

A. Comparison of POD rules to current work rules

In Exhibit III.1, we summarize the differences between the POD rules and the current work rules, following this with a more in-depth summary of the current rules and how the new POD rules attempt to simplify existing provisions. This context is especially important because some SSDI beneficiaries might not benefit from the benefit offset under the POD rules, which could directly affect their interest (and hence, recruitment).

Exhibit III.1. Comparison of POD rules to current rules



1. Summary of current rules

By statute, to qualify for SSDI benefits, an individual must be unable to engage in SGA. Given this definition, earnings above SGA are evidence that the beneficiary is able to work and therefore is no longer eligible for the program.

Consistent with this logic, SSDI beneficiaries earning more than the SGA level have their full benefit suspended if their earnings reach or exceed the SGA level for more than 12 months, though SSA has special work provisions to test work without losing benefits. These provisions include a nine-month TWP and three Grace Period months. A beneficiary immediately enters the Extended Period of Eligibility after completing the TWP. Through the end of the 36th month of Extended Period of Eligibility, SSA will reinstate benefits if earnings fall below the SGA level (the "re-entitlement" period).

Beneficiaries will face termination if earnings exceed the SGA level (the "cash cliff") after the re-entitlement period ends and the beneficiary has used all Grace Period months. Otherwise, benefits continue in full.

2. POD rules and associated services

POD includes work rule provisions, administrative processes to apply these rules, and benefits counseling (Exhibit III.2). The new work rule provisions attempt to simplify existing rules where beneficiaries had to track whether they had completed a TWP. This simplification also includes eliminating periods to test work (the TWP and Grace Period). Under the POD rules, SSA will use the new offset formula to adjust the monthly benefit based on monthly earnings. Benefits reduce $1 for each $2 increase in earnings above the POD threshold (of $850 in 2018). The new rules for reporting earnings should reduce the chance of reporting errors and overpayments if subjects report these issues to SSA in a timely fashion. Throughout the demonstration, POD subjects will face the same work rules until they reach full offset, the point when the subject's earnings cause SSA to reduce their benefits to zero. At this point, the POD rules differ for T1 and T2 subjects. T1 and T2 subjects both go into benefit suspension when benefits are fully offset, though T2 subjects face termination for 12 months of continuously receiving zero benefits. Finally, to support these new rules, the implementation team will provide benefits counseling and associated services to T1 and T2 subjects to make sure they understand the new rules and processes.

An important provision of POD is that T1 and T2 subjects can revert to current rules voluntarily at any point, which is specified as part of the requirement that no individual can be required to participate in the demonstration and has important implications for the evaluation. As part of our participation analyses, discussed in more detail in Chapter VII, we will closely monitor the number of T1 and T2 subjects who revert to current rules.

The Abt implementation team will provide benefits counseling to T1 and T2 subjects to explain the new rules and help them report earnings to SSA in a timely fashion to support the administration of the offset. The Abt implementation team has documented the service providers and the full set of intervention services and related supports in POD, designed to replicate the type of counseling that exists under current rules.[8]

[8] The Abt implementation team documented the intervention design services in an unpublished report to SSA (Abt Associates 2017a).

Exhibit III.2. Summary of POD provisions

New provision	Description
Simplified work rules	The POD rules eliminate the TWP and Grace Period. They also eliminate the need for a work continuing disability review during POD participation. As under current rules, SSA expects subjects to submit paystubs monthly.
Benefit offset	SSA will reduce monthly benefits by $1 for every $2 of monthly earnings more than the higher of the following: (1) the POD threshold amount and (2) total monthly itemized Impairment-Related Work Expenses. As under current rules, the POD threshold amount might change from year to year. If a subject has allowable Impairment-Related Work Expenses that are more than the POD threshold amount in a month, the benefit adjustment for that month will only apply to earnings above the Impairment-Related Work Expenses amount up to the current SGA level.
Benefits counseling	The implementation team will provide benefits counseling to POD subjects to make sure they understand the new rules.
Opt out to current rules	Subjects can voluntarily revert to current rules.

3. Implications of POD rules for beneficiaries

An important feature of the POD rules is that they will not always benefit all beneficiaries (Wiseman 2016). Specifically, some beneficiary subgroups might be worse off under the POD rules than under current law. This issue is important in understanding who might participate in the recruitment activities described below.

In general, the POD rules are favorable when a T1 or T2 subject has earnings above the current SGA amount, has few or no Impairment-Related Work Expenses, and has used up the Grace Period. Under current law, beneficiaries with earnings greater than SGA following the Grace Period will fall off the cash cliff and have zero SSDI benefits due. Conversely, under the POD rules, SSA would adjust benefits using the new offset rules by half of the amount of earnings above the monthly POD threshold amount.

However, there are cases when SSDI beneficiaries would have lower income under the POD rules. For example, under the POD rules, beneficiaries' benefits will be reduced when they earn more than the POD threshold amount, whereas under current rules, beneficiaries do not lose any benefits during the TWP and Grace Period. Additionally, beneficiaries with high Impairment-Related Work Expenses are less likely to benefit from the POD benefit offset. In Chapter II and Appendix, we describe other possible individualized scenarios where a beneficiary might not benefit from POD and their implications for the evaluation.

B. Recruitment and enrollment

As a starting point for recruitment, we will identify a population eligible for participation and then randomly assign the enrolled beneficiaries into the three study groups. Below, we describe how SSA and Abt worked together to identify the eligible sample. We then describe our approach to randomly assigning this eligible sample into the demonstration using both direct mailings to beneficiaries and an indirect outreach approach to bring in key stakeholders, such as advocates and other service providers, who can help establish the legitimacy of POD and, hence, facilitate recruitment efforts.

1. Eligibility criteria

Using program data that include the majority of working-age beneficiaries within each catchment area (Exhibit III.3), SSA will identify beneficiaries who meet certain eligibility requirements. These eligible beneficiaries are the potential subjects for POD. The general eligibility criteria include working-age beneficiaries (defined generally as ages 20 to 58 at the time of enrollment) who are either receiving SSDI benefits or whose benefits are suspended due to work at the time of study enrollment. Several other eligibility requirements cover special circumstances where beneficiaries could potentially face suspension or termination.[9] Additionally, SSA will exclude beneficiaries who are or have been in other SSA demonstration projects, such as BOND, to avoid conflating the POD findings with other SSA demonstration projects. Hock et al. (2017) lists the specific administrative criteria that SSA uses to pull the initial file of eligible beneficiaries for POD (for example, information about how SSA uses its program data to identify eligible beneficiaries, such as "primary" beneficiaries).

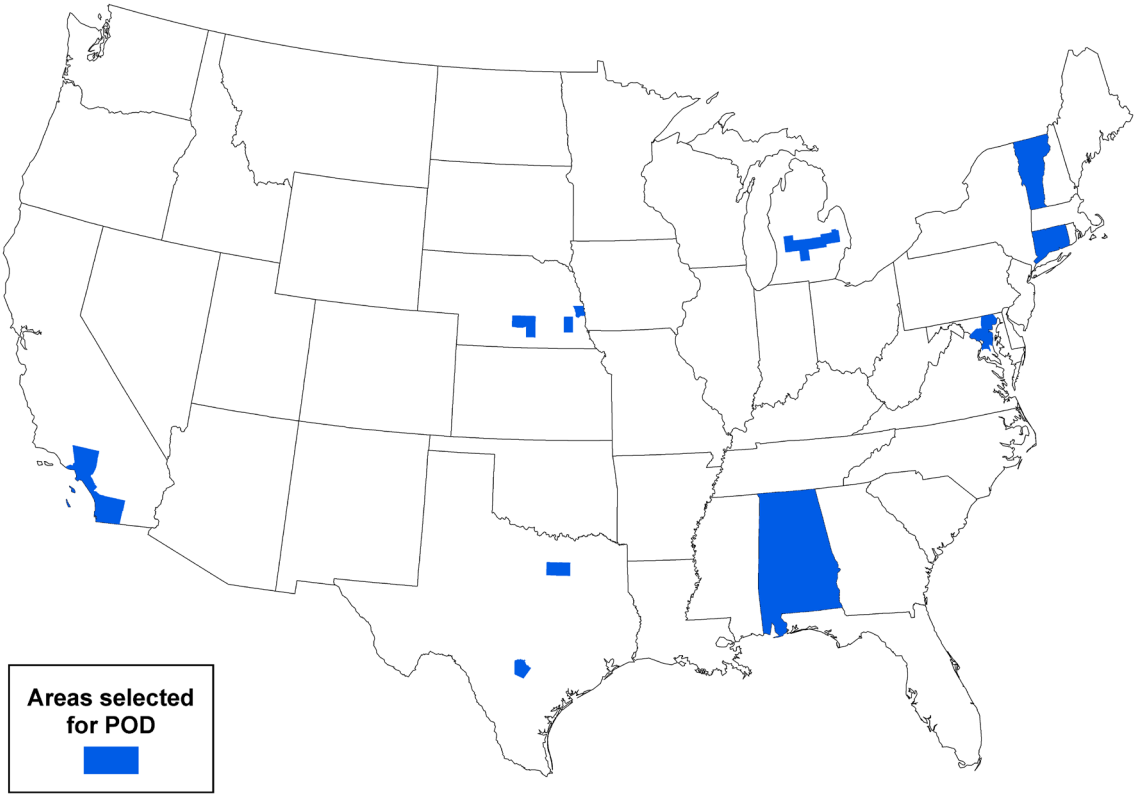
2. POD catchment areas

Abt and SSA identified eight states to include in POD based on the following criteria: (1) having sufficient numbers of SSDI beneficiaries to meet POD's target enrollment levels, (2) featuring a diverse range of beneficiary- and state-level characteristics, and (3) having VR entities willing and able to implement the demonstration design. Abt then engaged these states' VR agencies (or in some cases, VR regional offices, depending on the state's organizational structure) to identify catchment areas within the states where the demonstration will be implemented, using similar criteria as those for selecting the eight states.

Abt and SSA selected POD catchment areas purposively to cover different regions of the country, local labor markets, a mix of urban and rural areas, and a range of beneficiary characteristics (Exhibit III.3). Hence, by design, the study's recruitment pool will capture a broad range of local economic conditions and other contextual factors, though the areas are not statistically representative of the United States. POD includes the entire states of Alabama, Connecticut, and Vermont and subsets of counties in California, Maryland, Michigan, Nebraska, and Texas (Exhibit III.3). In states where a subset of counties was selected, POD catchment areas vary in urbanicity and population density. However, those counties tend to have relatively more cities and larger beneficiary populations than other counties not chosen for the demonstration because of the higher costs of serving beneficiaries in sparsely populated areas.

[9] For example, beneficiaries who have a pending work-related Continuing Disability Review are not included among those eligible for POD.

Exhibit III.3. POD catchment areas



Note: Areas originally included in the demonstration are shaded. The entire states of Alabama, Connecticut, and Vermont are included. Subsets of counties in the other five states are included. State-level maps are in Appendix B. Additional areas were added after the recruitment pilot (see addendum).

3. Direct and indirect recruitment outreach

Recruitment and enrollment will rely on the following two key factors:

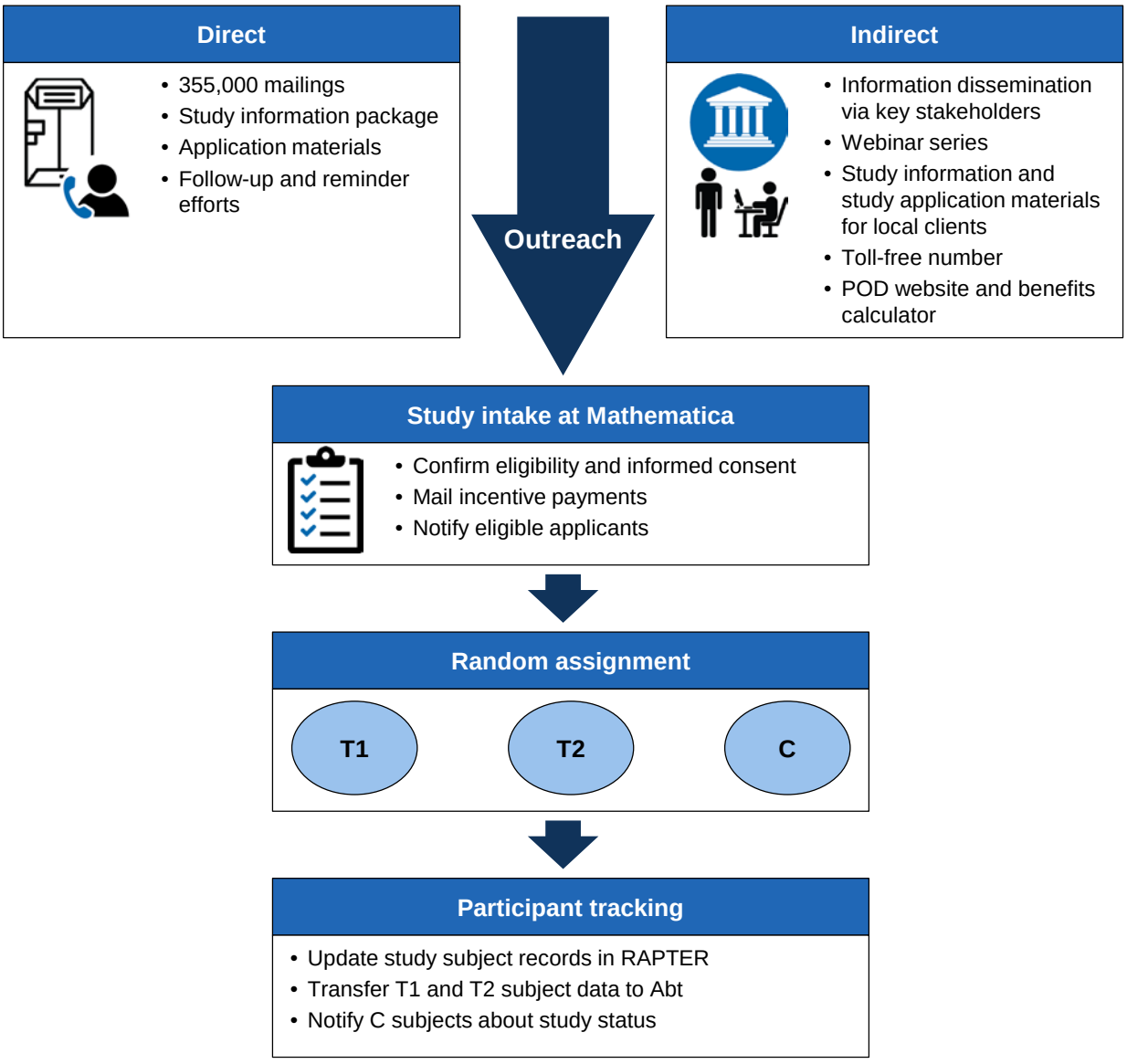
- Direct outreach to prospective subjects
- Indirect outreach via trusted stakeholders—organizations commonly engaged with beneficiaries, especially around employment issues

Key elements of both direct and indirect outreach are shown in Exhibit III.4 for the pilot stage, and we will initiate them in tandem so that beneficiaries receive coordinated information. Direct outreach will take the form of mailings to all eligible beneficiaries identified by SSA in these sites (approximately 355,000 beneficiaries). In these mailings, we will explain POD and ask the beneficiaries to return completed baseline surveys and consent forms. At the same time, our indirect outreach efforts to key stakeholders will establish the legitimacy of POD in each implementation area, a critical component of successful recruiting (Derr et al. 2015).

We will recruit volunteers from the POD-eligible list that SSA provides each month, sending waves of mailings to a representative random sample of beneficiaries. Monthly sampling waves for each state will be based on "replicate" groups of beneficiaries randomly selected from the eligible list. Within each state, we will use implicit stratification to increase the degree to

which each replicate is representative of the POD-eligible list. We will first create strata for each of three primary diagnoses (injuries, neoplasms, and severe visual impairments) that we expect to be particularly distinctive in terms of the attractiveness of the POD offset. Among the remaining beneficiaries, we will create strata based on combinations of age, duration receiving SSDI benefits, and whether they had TWP-level earnings between January 2014 and October 2017. Beneficiaries may voluntarily enroll in POD as long as they meet the eligibility criteria and return completed consent and baseline questionnaires. The evaluation team will screen out any volunteers who decline to consent, have moved out of the demonstration catchment area, send an incomplete baseline questionnaire, or return the forms after recruitment has ended for the state.

Exhibit III.4. POD outreach and enrollment



To enroll in POD, eligible beneficiaries must complete and submit the aforementioned baseline survey and consent form. We will verify continued eligibility; check that the consent form is complete; and, based on responses to the consent form and two short questions in the survey, determine whether the beneficiary understands the parameters of the benefit offset and voluntary nature of the demonstration. Finally, we will randomly assign beneficiaries who complete all these steps to one of three study groups (T1, T2, or C); send respondent payments to every beneficiary who applied; and, in coordination with Abt, notify study subjects of their enrollment in POD. For more details on the specific random assignment processes, see Hock et al. (2017).

C. Monitoring recruitment

Our recruitment plans build off the experience of recruiting volunteers for BOND, and we will use the pilot period to refine our assumptions and recruiting approach. Though BOND provides a broad indication of the level of interest in a benefit offset, the evaluation team expected a lower volunteer rate for POD. Unlike the BOND rules, the work rules for POD could leave some beneficiaries worse off than current law, and the Bipartisan Budget Act (Section 822) required SSA to obtain written consent from beneficiaries who want to enroll in POD. In the remainder of this section, we discuss our assumptions for volunteer rates, the pilot recruitment period used to test those assumptions, and possible refinements.

1. Assumptions for POD volunteer rates

We plan to enroll at least 9,000 beneficiaries in POD after sending mailings to over 355,000 beneficiaries, although the demonstration can accommodate up to 15,000 study enrollees. To meet the minimum enrollment target of 9,000, we would need a volunteer rate of approximately 2.5 percent, and the initial findings from the pilot recruitment period are broadly consistent with such a volunteer rate. During the remainder of the recruitment period, we will continue refining our assumptions about take-up rates—overall and by subgroup—and use this information to update our sampling approach, as needed.

We expect the selected catchment areas shown in Exhibit III.3 will include more than enough eligible beneficiaries to meet POD's study enrollment target, and POD could accommodate larger enrollment (up to 15,000) if refinements to the recruitment process succeed in increasing the expected enrollment yield. Exhibit III.5 shows how the 355,460 eligible beneficiaries in the POD catchment areas are divided across states and our projections of how they would be apportioned in the sample if the final enrollment total were the minimum of 9,000. If the enrollment yield increases overall or for individual states, we will consult with SSA and Abt about whether and how to adapt recruitment, such as by accommodating a greater enrollment or by targeting recruitment to a subset of beneficiaries or catchment areas.

The study subjects from these states will not be statistically representative of *all* SSDI beneficiaries for two reasons. First, the eight selected states—and the counties within those states—do not represent a nationally representative sample of population of beneficiaries. Second, the demonstration includes volunteers who self-select for services. We will target our direct outreach efforts on a representative sample of eligible beneficiaries living in POD catchment areas, but only those who are interested in participating will reply.

Exhibit III.5 also illustrates that our sample sizes vary considerably across the eight states. Three of the states—Alabama, California, and Texas—have substantially larger samples than the other five states, and Nebraska and Vermont have substantially smaller samples. Some of this variation is due to the different sizes of the state beneficiary populations, but it also reflects differences in the number of counties included within the states. In particular, Alabama's catchment area spans the whole state, which is why the number of Alabama beneficiaries is close to the expectations for California and Texas, where POD covers a subset of counties. Conversely, Nebraska has a larger population than Vermont but only covers a few counties, so the two states have similarly small expected sample sizes.

Exhibit III.5. Summary of POD direct mailings

State	SSDI beneficiaries in selected catchment areas meeting eligibility criteria	Projected number of potential POD study subjects[a,b]	Proportion of POD enrollment target[b] (%)
Alabama	68,656	1,738	19.3
California	99,634	2,523	28.0
Connecticut	38,532	976	10.8
Maryland	40,417	1,023	11.4
Michigan	22,152	561	6.2
Nebraska	11,997	304	3.4
Texas	67,563	1,711	19.0
Vermont	6,509	165	1.8
Total	**355,460**	**9,000**	**100.0**

Source: SSA program data as of April 4, 2018 with preliminary pilot period information.

Note: We will update these projections upon receiving updated information from SSA about the number of eligible beneficiaries in POD states. See the previous section for POD eligibility criteria.

[a]We projected the number of POD study subjects to be slightly above 2.5 percent of the number of eligible beneficiaries, although in practice the study volunteer rates will likely vary from state to state.

[b]We have shown the total number of potentially eligible subjects in the table, with the understanding that we will later work with SSA and Abt to avoid substantially changing enrollment without advance agreement.

2. Testing assumptions during the pilot

We will implement a two-month pilot recruitment period to learn about recruitment yields and experimentally test variants of our outreach approach (Hock et al. 2017). We will use this information to improve our outreach efforts over the remaining 10 months of the recruitment period to better support SSA's aims for this demonstration and evaluation. For example, in consultation with SSA, we might adapt our targeting of reminder and encouragement efforts to improve yields or achieve more diversity in the types of beneficiaries enrolled in the demonstration, as well as adjust some core parameters of our recruitment strategies. Abt and its state implementation partners can also use the pilot period to test their operations and interactions with beneficiaries and, as necessary, adjust service delivery protocols before full rollout. Whatever changes occur, subjects enrolled during the pilot period will be full-fledged POD enrollees in every respect.

Initial pilot (January 2018 to February 2018). During the pilot period, we will target a sample of beneficiaries randomly selected from the list of eligible beneficiaries in the POD catchment areas. We will send 30,000 mailings during the initial two-month period. We will select the pilot sample in the same fashion as described for the main sample. Based on our assumed yield rate assumptions, this initial mailings during the pilot will translate to 900 study subjects. We plan to include all pilot enrollees in the evaluation's analyses, which is consistent with the approach used for the Accelerated Benefits demonstration and the second stage of BOND (Michalopoulos et al. 2011; Gubits et al 2013). The primary reason to include pilot subjects is that SSA will incur costs in serving these beneficiaries and, unless there are substantive differences in their services or characteristics, they can be studied in a pooled sample with later enrollees. Alternatively, if we do find such differences, we will consider using analysis models to control for the factors that varied between the groups. We might also consider presenting results for pilot subjects as a separate subgroup from subjects enrolled later.

Full rollout (March 2018 to December 2018). Following completion of the pilot, we will recruit the remaining beneficiaries in the POD catchment areas, making changes based on the results of the pilot. During the full rollout, we will enroll at least 8,100 additional subjects, assuming the pilot meets its recruitment goal. We have planned to mail all of the remaining eligible beneficiaries in the POD catchment areas to reach the target of at least 9,000 subjects. We will coordinate these recruitment efforts to draw a steady flow of subjects into POD throughout 2018, to help Abt manage its workload.

3. Possible modifications for full rollout to achieve sample targets

We will analyze data on recruitment yields during the pilot period along with the results of the recruitment experiment to assess potential adaptations to the core recruitment methods. We will make any necessary adaptations to the recruitment strategy to reach the sample goal of enrolling at least 9,000 subjects. Our goal is to identify potential adaptations that balance both resource considerations, such as cost-effectively achieving a yield rate that meets the enrollment target, and evaluation considerations, such as drawing in study subjects whose characteristics more closely resemble the broader population. We will work with SSA to weigh these trade-offs and establish how best to fine-tune the recruitment process for the full rollout. Our plan is to target the adaptations to reach this goal based on the budget, while also summarizing the evaluation implications. In this assessment, we will pay particular attention to options that are likely to help meet the study enrollment target while including a sample of POD subjects who cover a wide range of work trajectories after enrollment. Accordingly, we will develop recommended adaptations to the recruitment activities and brief SSA on options, proceeding with the adaptions once approved.

D. Implications for recruitment and evaluation

We expect the study sample will reflect a broad array of beneficiary characteristics and outcomes, but this might not allow us to reliably predict the experiences of beneficiaries who are not in POD's catchment areas or choose not to apply. There are several key implications of the recruitment plan and resulting study sample that we will need to account for when we interpret the impact estimates:

- **The eight selected states and the counties within those states are not a nationally representative sample of the population of beneficiaries.** The catchment areas are purposively selected based on implementation needs, though in selecting the states Abt also considered the value of having states from different regions and with different sizes and economic conditions.
- **The eligible beneficiaries who enroll are volunteers who are self-selecting into the demonstration and are not representative of the population of beneficiaries.** Individuals who choose to enroll in the study are likely to be fundamentally different from those who do not. For example, some beneficiaries will not volunteer for POD because they recognize the new rules will not be of benefit. Some of these fundamental differences might be measurable, such as having stronger work histories. Other differences might not be observable, such as having stronger motivation to earn enough to not require benefits.
- **The benefit offset we are evaluating includes the option for study subjects to revert to current rules.** Once the demonstration starts, some T1 and T2 subjects might revert to current rules if they fall into one of the aforementioned categories of beneficiaries who are better off under current rules than the POD rules. We anticipate that the incentive to revert to current rules will be much stronger for T2 subjects given the provision that T1 subjects will not have their benefits terminated for excess earnings. Tracking withdrawal rates and the characteristics of study subjects who withdraw will be an important component in the participation analysis (as discussed in Chapter VII).
- **Monitoring recruitment to assess progress toward sample targets will be crucial.** This will be an especially important objective of the pilot phase. The pilot may also reveal lessons about how to target outreach to be cost effective and meet evaluation objectives, and such targeting could further change the sample's characteristics.

Nonetheless, the study's random assignment design will provide rigorous impact estimates for study subjects, as discussed in more detail in Chapter VII. Additionally, special topics reports will consider alternative means of accounting for self-selection and extrapolating to broader populations.

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

IV. QUANTITATIVE DATA

The evaluation will draw on the following three distinct types of data that feed into the planned analyses described in the subsequent chapters:

- ***Administrative data.*** Similar to previous evaluations of SSA demonstrations, such as BOND and Youth Transition Demonstration, the POD evaluation will include program and earnings data that inform all of our planned analyses. These data include (1) SSA program data, which provide information on beneficiary characteristics and several programmatic outcomes (for example, SSDI benefits); (2) data on annual earnings, derived from SSA's Master Earnings File; and (3) VR program data, which provide information on VR programmatic outcomes. We plan to use these data to measure characteristics and outcomes throughout the demonstration. Additionally, SSA can use these data to assess long-term evaluation outcomes after the demonstration ends.
- ***Management information system (MIS) data.*** Our evaluation team and the implementation team have separate MIS data sources, both of which we will use to support evaluation activities. The Mathematica system, Random Assignment Participant Tracking Enrollment and Reporting (RAPTER), will allow our team to implement, manage, and report on the intake and random assignment process for all POD subjects. We will use these data to summarize findings about recruitment, withdrawals, eligibility status changes, and other status changes in the participation and process analyses. Abt's Implementation Data System will allow our team to track service delivery to POD treatment subjects. We plan to use these data to document POD processes and the intensity of participation among POD treatment subjects.
- ***Surveys.*** We will administer three surveys with beneficiaries who enroll in POD: a baseline survey and two follow-up surveys, at Year 1 and Year 2 post enrollment, respectively. The baseline and two follow-up surveys provide information that is not available in administrative data. The characteristics from the baseline survey will support the participation analysis to facilitate comparisons between treatment and control subjects. Additionally, we plan to use these baseline data as control variables in the impact analysis. The follow-up surveys will allow our team to assess impacts of POD on outcomes not included in the administrative data described above, such as health, job characteristics, and metrics that are relevant to the benefit-cost analysis (for example, participation in other government support programs). Additionally, the follow-up surveys include relevant information to the process study about perception of POD supports among POD treatment subjects.

Each of these data sources includes associated "data lags," which have direct implications for our planned time frames in reporting findings to SSA (see Chapter X for a summary of the reports). Specifically, for each data source, we have to obtain the information, create an analytic file, and then analyze findings for each of the reports. Some data sources include monthly updates, such as SSA program data and MIS data, and hence have relatively "short" data lags. Thus, we can potentially use these types of files to produce analytic information within the same calendar year. However, other data sources have less frequent updates, such as annual earnings

updates from SSA. The reports using these types of data tend to have a "moderate" data lag, where it takes 18 to 24 months to update.

In our tables, we note "file reporting", which defines when the files are updated within each system. As part of the evaluation, we will make data requests in preparation for each evaluation report. For example, some data, such as SSA program data and Abt Implementation Data System, are updated monthly.[10] For the evaluation, we will make a one-time request for the files to coincide with the needs of the evaluation (as opposed to building a comprehensive evaluation data system that is updated in real-time).

In the remainder of this chapter, we describe each of the data sources, including specific data elements that we plan to integrate into the analysis and the anticipated data lags (described as "short" versus "moderate"). Our summary of data elements illustrates how we plan to integrate these data sources into each of the analyses described in Chapters VI (process analysis), VII (participation analysis), VIII (impact analysis), and IX (benefit-cost analysis). Our summary of data lags provides context for the types of information we can include in each report to SSA, as discussed in Chapter X.

A. Administrative data sources

We will work with SSA and state VR agencies to integrate information from administrative data systems, enabling us to track beneficiary characteristics and outcomes (Exhibit IV.1). These data include SSA program files that contain information about beneficiary characteristics at random assignment and longitudinal information about program and earnings outcomes. Our team will work with SSA to access these data following the security procedures developed for the demonstration specifically to protect against unauthorized access to data and inappropriate use of data by authorized users.

1. SSA program files

SSA maintains several program files that it regularly updates in administering Supplemental Security Income (SSI) and SSDI payments and monitoring eligibility. As in previous evaluations, SSA staff will obtain program information from several internal administrative systems that we will develop into analytic files for the analysis.[11] SSA staff will work with Mathematica staff to provide access to these data and produce extracts into a usable research file to support the POD evaluation.

[10] A detailed description of Abt's Implementation Data System is below (see IV.B.2).

[11] SSA will draw on several internal files to produce necessary data elements, including the Master Beneficiary Record (for example, monthly SSDI payment status); Supplemental Security Records (for example, monthly SSI payment status); NUMIDENT (date of death); Disability Control File (for example, Impairment-Related Work Expenses). We will use the Disability Analysis File, which contains much of the program data we need, but is not updated as timely as required to meet the reporting schedule. SSA makes the technical documentation developed under the Disability Analysis File contract available broadly to SSA program data users; the documentation includes code, information on source files, mainframe processing, and related technical procedures we can draw upon for POD.

The SSA program files include detailed information about beneficiary demographic, impairment, and program characteristics that will be useful in our assessment of POD recruitment and participation. Specifically, these data include information about age, sex, impairment, and historical program information that our team plans to use to construct profiles of POD subjects. We will also use this information to examine the characteristics of beneficiaries who did and did not participate.

The program files also include longitudinal benefits information on several outcomes, which are especially useful for tracking impacts. Examples of these variables include SSDI and SSI program status, benefits paid, overpayments, and use of SSA work incentives. We used each of these variables in the previous BOND evaluation reports. Hence, we expect to continue using similar constructs here to provide transparency in reporting and, eventually, facilitate comparisons of outcomes between SSA evaluation reports, which SSA may desire in communicating findings to outside agencies and policymakers.

One notable feature of the SSA program data is the retroactive adjustment process, which has important implications for interpreting benefit payment and overpayment outcomes in each report. For example, all existing BOND evaluation reports presented impacts on benefits paid *in* a given calendar year rather than benefits paid *for* that year (see Wittenburg et al. 2015). The benefits-paid-in measure reflects all current benefit payments to the beneficiary but does not initially reflect retroactive payment adjustments or improper past payments recovered during that year. SSA eventually updates program data to reflect retroactive payment adjustments. Conversely, the benefits-paid-for measure reflects benefit payments after all retroactive payment adjustments. In the later years of the BOND evaluation, the BOND evaluation team estimated impacts using the benefits-paid-for measure for the earlier years of the report (Hoffman et al. 2017).

We plan to follow a similar format as that in the BOND reports in depicting benefits to account for retroactive updates. Specifically, we will add caveats that our impact findings in earlier reports for benefits and overpayments could change later in POD. We will note that benefits paid in a particular year could differ from benefits paid for that year. In the POD final report, we will assess whether any of our earlier impacts differ between the benefits-paid-in versus benefits-paid-for measures, as well as produce impact estimates for overpayments.

In general, the analytic lags associated with program files are relatively short given the SSA monthly data structure. For example, we can generally work with SSA to quickly obtain program data because its program files have monthly reports for benefits paid in a year and eligibility. As noted, there is a lengthier period necessary to observe benefits paid for, given the retroactive adjustment period.

Exhibit IV.1. Summary of selected variables from administrative files

Potential variables	SSA program files	Annual earnings data	State VR data
Beneficiary characteristics			
Demographic characteristics (age, sex)	X		
Diagnosis, impairment status	X		
Historical program information	X		
Historical earnings information		X	
Outcomes: Program participation and earnings			
SSDI eligibility, SSI eligibility, program exits	X		
Benefit payments for, benefit payments in year, overpayments	X		
Use of work incentives (Impairment-Related Work Expenses, Ticket to Work)	X		
Earnings above annualized POD threshold level	X	X	
Annual earnings		X	
VR application date			X
Types of VR services received			X
VR closure status and reason			X
VR employment at closure variables (hours, wages)			X
File reporting for research purposes	**Monthly**	**Annually**	**Monthly**

Notes: The SSA program files include information from the Master Beneficiary Record, Supplemental Security Records, NUMIDENT, and Disability Control File. The annual earnings data include information from the Master Earnings File. Finally, the state VR data include state data from VR agencies.

2. Annual IRS earnings data

We will work with SSA to obtain annual earnings information for POD subjects from the Master Earnings File (hereafter referred to as Internal Revenue Service (IRS) earnings data), which provides information on annual employment and earnings for all POD subjects. Our employment and earnings measures will represent all earnings reported to the IRS. The only earnings excluded are any earnings that people do not include in their filings to the IRS. We plan to use these data to construct several measures of employment, such as any reported earnings and earnings above certain thresholds that mark key milestones for POD outcomes, such as earnings above the annualized POD threshold amount.

The IRS updates earnings daily for the Master Earnings File and, based on Tax Year 2016 experience, 99 percent of tax records will have updates by June of the following calendar year. SSA considers the earnings data fully updated by the following February (14 months after the previous year). For example, the earliest period for which we can report, say, 2018 annual earnings outcomes in the evaluation, is July 2019 (99 percent update), and we can assume those same earnings would be fully updated by February 2020.

Another issue that adds to the time lags is that access to the data are restricted. Specifically, based on authority granted in the Internal Revenue Code, only qualified SSA staff may access the IRS earnings data; contractors may not use the data directly. Hence, our team must submit

programs through SSA and ensure there is enough time for processing before reporting findings publicly. Qualified SSA staff access the data, submit programs developed by our team to estimate impacts, review output to ensure that it complies with privacy requirements, and then transmit the output to the evaluation team. In Chapter X, we depict the annual outcomes we can include in each report based on this experience.

3. State VR data

Our final source of administrative information includes monthly data from VR case services from state VR agencies, which will provide regularly updated information about VR processes and outcomes.[12] Our team has current experience in using these data for an ongoing SSA demonstration (PROMISE), which we plan to use as a template for the POD evaluation in generating VR outcomes. Based on the PROMISE experience, we expect the VR data elements to include whether the beneficiary applied for VR services during the course of the intervention, the duration of receipt of VR services, the types of services received (such as job placement or college training services), and closure outcomes (including exiting from VR with employment). Also, based on the PROMISE experience, we expect the state VR program data for POD will have relatively minor data lags (similar to the SSA program data), enabling us to obtain updated data for the impact reports.

B. MIS data

We will use information from our own recruitment data system, RAPTER, and work with Abt to obtain service information from its Implementation Data System. The RAPTER data include recruitment data for treatment and control subjects. Our survey team will load beneficiary characteristics from SSA program data into RAPTER to develop the address lists for the mailings and manage the overall recruitment process. Hence, these data include information about beneficiary characteristics, information about the random assignment process, and information about eligibility status changes or withdrawal from the POD rules (and reversion to current rules), as well as participation in the follow-up surveys. Abt plans to use the Implementation Data System to manage its delivery of POD services. Hence, we can request extracts of these data from Abt to examine service use, such as questions about POD and corresponding use of work incentive supports, by treatment subjects.

Similar to the SSA benefits data, the data in RAPTER and the Implementation Data System extracts have relatively limited time lags because both Abt and Mathematica will regularly update these systems daily. During the recruitment phase, we will request data extracts and present statistics monthly to coincide with our requests for SSA program data noted above. This

[12] As outlined in Deliverable 6.2, we will work with SSA in developing a strategy for acquiring VR program data, which will vary depending on the implementing entity in each of the demonstration states. For each of the four state VR agencies with which Abt has subcontracted to implement POD (Alabama, Connecticut, Maryland, and Vermont), Abt will include data use agreements when establishing subcontracts. These agreements will specify the data the VR agencies will submit to Abt to support the evaluation of POD. Abt will securely transfer the VR program data to SSA according to the agreed-upon data submission schedule. In the remaining four states (California, Michigan, Nebraska, and Texas), we expect that SSA will establish separate data use agreements with VR agencies given that the WIPA provider (and not the VR agency) will implement POD.

reporting provides a natural setup for the evaluation to depict service delivery processes and program outcomes across the same time periods.

1. Mathematica's RAPTER system

Our evaluation team will use RAPTER data to depict information about the recruitment efforts for the participation analysis (top panel of Exhibit IV.2). The RAPTER data include information about the characteristics of subjects from SSA program data; direct outreach (number of mailings sent); recruitment and enrollment (for example, complete recruitment packets); and random assignment status (T1, T2, or C). For example, we will use the data from this system to compare the characteristics of all of the eligible beneficiaries in the POD solicitation pool (see Chapter II) to POD subjects. We will also use these data to examine the characteristics of POD subjects, especially in establishing baseline equivalence of the three study groups, to assess whether random assignment worked as envisioned.

2. Abt's Implementation Data System

We will use Abt's Implementation Data System to support the process and participation analyses (bottom panel of Exhibit IV.2). Specifically, for the process analysis we will use the Implementation Data System to examine provision of counseling, how POD states and the POD support units facilitate and manage monthly reporting of earnings and Impairment-Related Work Expenses, assess whether the intervention is being implemented as intended, and identify which aspects of POD have been implemented with low fidelity. The participation analysis will use these data to examine the incidence and frequency of reporting monthly earnings and Impairment-Related Work Expenses as well as benefit offset use, engagement in counseling services, and patterns of participation in the demonstration, including an analysis of those subjects who withdraw from POD.

3. SSA's POD Automated System

We will use data from SSA's POD Automated System to understand POD benefit offset use, SSA's disposition for Impairment Related Work Expenses claimed by treatment subjects using the POD offset, and also overpayments and underpayments that result after SSA annually reconciles benefits to identify the correct amount of SSDI benefits that should have been paid to each treatment subject under the POD offset rules during the prior calendar year. The POD Automated System is an SSA internal computer system that accepts data files from Abt's Implementation Data System containing earnings and Impairment Related Work Expenses information that will be used to administer the POD benefit offset. The POD Automated System will calculate the POD benefit offset amount and will capture information from each treatment subject's Master Beneficiary Record to determine the correct monthly benefit amount. The POD Automated System will create a Manual Adjustment Credit and Award Process action that updates the Master Beneficiary Record in an overnight batch process. When the POD Automated System receives the earnings report from Abt's Implementation Data System, the data system will calculate the offset amount, retrieve information from the Master Beneficiary Record, and will then determine if the case can be processed automatically. If the case can be processed automatically, the POD Automated System will adjust the benefit payment based on the treatment subject's earnings and Impairment Related Work Expenses reported to Abt and the treatment subject's previous months monthly benefit amount record in the Master Beneficiary

Record. The POD Automated System will then send a notice to the treatment subject of the benefit change. If the case cannot be processed automatically through the POD Automated System, the system will generate an edit code and the processing limitation, at which point specialized units within the Processing Center of jurisdiction receives and resolves the edit code. The Processing Centers will work the case manually and will update the POD Automated System database with the offset determination. After the close of the calendar year, the POD Automated System will determine whether underpayments or overpayments are applicable for the prior year by running a match against the Master Earnings File earnings and the monthly earnings from the treatment subjects reported to Abt.

C. Surveys of beneficiaries

Mathematica will collect three surveys for POD. First, we will collect a baseline survey as part of the effort to solicit volunteers to participate in POD. The baseline survey will be short, given that we can extract several demographic and impairment characteristics from the SSA program data noted above. We will also conduct two follow-up surveys to collect information about outcomes not included in the SSA program data.

The time lag associated with each of the surveys is moderate because we have to wait until the final interviews are complete and the survey team can clean the data for research purposes. As described in Chapter II, POD includes a rolling random assignment period that covers approximately 12 months, and the follow-up surveys mirror that setup in their one- and two-year follow-up periods. Hence, below, we depict the time frame for each survey that corresponds with recruitment and then the follow-up period.

1. Baseline survey

To participate in POD, beneficiaries must complete a 20-minute baseline survey and return it along with a signed consent form (Exhibit IV.3). As documented in Hock et al. (2017), we will mail the baseline survey and a consent form as part of the POD recruitment packet. Beneficiaries must complete both the questionnaire and consent form before random assignment. We will use the survey to collect baseline information for the evaluation that is not readily available in program data.

We plan to use the baseline survey data for the participation and impact analyses. For example, for the participation analyses we plan to use the baseline survey data to facilitate baseline equivalence comparisons of the study group that are not included in the RAPTER data (from SSA program records), such as current work status, education, and income. For the impact analysis, we will also use this information to form subgroups or adjust for baseline characteristics when estimating the impacts of POD.

Exhibit IV.2. Data sources from Mathematica, SSA, and Abt

Data source	Key data categories
Mathematica's RAPTER	
Direct outreach	Number of outreach mailings sent Number of outreach mailing returned as invalid address or recipient Number of reminder phone calls by state or county region Number of reminder calls that connect, do not connect, or are invalid Percentage of SSDI beneficiaries reached who subsequently enrolled in POD
Recruitment and enrollment	Number of incoming calls to 1-800 number Percentage of enrollment packets returned with completed survey Percentage of enrollment packets returned with completed consent Percentage of enrollment packets returned without beneficiary volunteering Percentage of beneficiaries who required re-contact by Mathematica to complete consent form Percentage of enrollment target reached for each study group Status of each received packet (ineligible, consent not granted, quality control callback, awaiting consent) Respondent payment processing
Random assignment	Study group assigned (T1, T2, C) Demographic characteristics of eligible subjects
SSA's POD Automated System	
Administration of benefit offset	Earnings reporting (reminder letters/emails, dates reports submitted, reporting mode, monthly Impairment-Related Work Expenses amounts/type, earnings over POD threshold) Earnings record processing (processing time, follow-up and quality control review, percentage of T subjects known to be over the POD threshold with complete records submitted to SSA timely) POD benefit adjustment Annual benefits reconciliation (overpayments, underpayments) Benefit adjustment appeals
Abt's Implementation Data System	
Enrollment into and transition out of POD	Date, mode of initial contact with T1 and T2 subjects post random assignment Outreach attempts Number of subjects who withdraw from POD (by state or county region) Date and reason for withdrawal from POD
POD Work Incentives Counseling (WIC) and other supports	WIC staffing (providers, caseload size) Trainings delivered to WIC staff Remote service delivery Onboarding of new T subjects Development of benefit summary and analysis, work incentives plan Delivery of ongoing work incentives counseling Provision of technical assistance
File reporting for research purpose	Monthly

Note: RAPTER and Abt's Implementation Data System are updated daily when they are in active use. We will update RAPTER manually and less frequently when recruitment and survey fielding are inactive but when there could still be relevant updates due to withdrawals or other changes in status. We will request file updates for the MIS data to correspond with the update periods for the SSA program data.

Exhibit IV.3. Baseline and follow-up survey content

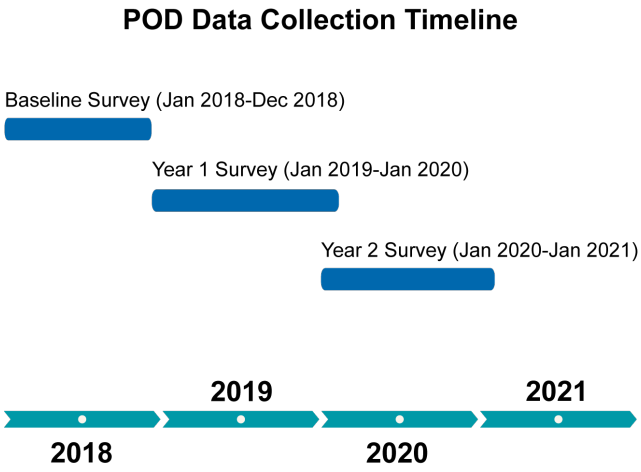
Survey Content	
Baseline	
Characteristics	Education—highest achieved
	Overall health status and use of health insurance
	Race and ethnicity
	Current living and housing arrangements
	Work limitations
Pre-POD employment, earnings & income	Date of most recent job, and whether looking for work in past four weeks (if not currently working)
	Earned at least $1,000 in any month over the past 12 months
	Likelihood of working in next 12 months
	Income—total household
	Currently working at job for pay
Pre-POD benefits & services	Job training experience
	Receipt of services from a benefit specialist or WIPA provider
Follow-up	
Employment	Any jobs for pay in past 12 months
	Hours worked, fringe benefits offered, and other details of current, main, or most recent (post-RA) job
	Work accommodations, job satisfaction, attitudes toward work and returning to work
	Job search activities
Earnings	Wages received at current, main, or most recent (post-RA) job
Benefit receipt	Understanding/attitudes toward the POD offset, termination of benefits
	Receipt of services from a benefit specialist or WIPA provider (post-RA)
	Satisfaction with POD offset, rules and services, reasons for withdrawing from the POD offset
	Job training experience and education since RA
Income	Income from workers' compensation, private disability insurance
	Income from other social programs, such as housing assistance, Supplemental Nutrition Assistance Program, Temporary Assistance for Needy
	Household income from all sources
Other	Physical and mental health status, hospitalization, current health insurance
Data collection end period for file reporting	• Baseline survey (final interview: December 2018) • Year 1 survey (final interview: January 2020) • Year 2 survey (final interview: January 2021)

2. Year 1 and Year 2 follow-up surveys

We designed the two follow-up surveys to capture the same content over one-year intervals. The one-year follow-ups provide SSA with annual updates of POD's progress. In the original proposal, SSA developed the time frames for these surveys to coincide with the development and delivery of the interim and final report (see Chapter X). The surveys include information on several outcomes not measured in the SSA program or MIS data. Specifically, both surveys

include information about short-term employment-related activities, training and education, receipt of and satisfaction with POD services, understanding and attitudes toward work incentives, health and functioning, total income, and other contextual variables (see Exhibit IV.3).

Other than the longer follow-up, the primary difference between the Year 1 and 2 surveys is that the Year 2 survey will include the full sample of POD subjects, whereas the Year 1 survey will include a 50 percent subsample of subjects. Specifically, we will conduct the Year 1 survey with a subsample of POD subjects (7,500 subjects) beginning in fall 2018, which will feed into our report on interim outcomes. We will conduct the Year 2 survey with the full sample of 15,000 POD subjects beginning in fall 2019, which will feed into the final evaluation report on longer-term outcomes.



The baseline and follow-up surveys will coincide with the period for enrollment outlined in Chapter II. Specifically, the baseline survey will have a 12-month field period, starting in January 2018 and continuing through December 2018. The Year 1 survey will occur one year following the baseline survey, starting in January 2019. It will include a rolling period that tracks along with the baseline survey and an additional one-month field period to collect all of the data from the final release. Hence, the full field period for the follow-up survey will be 13 months (January 2019 through January 2020). The Year 2 survey will occur two years following baseline and, similar to the Year 1 survey, will have a 13-month field period (January 2020 through January 2021).

We plan to use the follow-up survey in the process, impact, and benefit-cost analyses. For the process analysis, the follow-up survey includes information with which we can conduct simple descriptive tabulations to assess beneficiary satisfaction with POD services. For the impact analysis, we can use both follow-up surveys to assess impacts not included in the program data, such as health and total income. Finally, for the benefit-cost analysis, the follow-up survey includes measures of participation in other programs that we can use to quantify the benefits and costs of POD across several programs.

D. Evaluation implications

In Exhibit IV.4, we summarize each of the data sources described above by its planned use in each of the analysis and updates. From the administrative data, the monthly SSA program files will be a central source of information for the evaluation and included in each of the planned analyses. We will also include the annual earnings data from SSA and state VR program data to support the impact and benefit-cost analyses. We will include monthly information from the two MIS data sources (RAPTER and Abt's Implementation Data System) to support the process and

participation analyses. Finally, the surveys will include information not available in the administrative data to support all of the planned analyses.

Exhibit IV.4. Summary of quantitative data sources for evaluation

	Process analysis	Participation analysis	Impact analysis	Benefit-cost analysis	File updates
Administrative data					
SSA program data	X	X	X	X	Monthly
IRS earnings data			X	X	Annual
VR program data			X	X	Monthly
MIS data					
Mathematica's RAPTER (recruitment data)	X	X		X	Monthly
Abt's Implementation Data System (service data – *treatment subjects only*)	X	X		X	Monthly
Surveys					
Baseline surveys		X	X		Final Interview: December 2018
Year 1 follow-up survey	X		X	X	Final Interview: January 2020
Year 2 follow-up survey	X		X	X	Final Interview: January 2021

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

V. QUALITATIVE DATA

We will collect qualitative data to supplement our quantitative data collection to provide SSA with a comprehensive evaluation of POD. Our qualitative data collection will focus on domains that overlap with the quantitative data, but with greater emphasis on understanding how SSA and Abt are implementing the offset, barriers and facilitators to implementation, and in-depth perspectives from beneficiaries. Specifically, we will collect qualitative data from the following four groups of stakeholders:

- ***Implementation management and site staff.*** We will conduct interviews with Abt's implementation team to learn more about service delivery to help treatment subjects use the new offset and associated services. The implementation staff includes several entities: Abt's management that oversees the implementation activities, POD work incentives counselors (POD counselors), VR agency/WIPA managers supervising POD counselors in the POD sites, VR agency staff, WIPA providers, and technical assistance providers from Virginia Commonwealth University.
- ***POD treatment subjects.*** We will conduct semi-structured interviews with POD treatment subjects to learn more about their perspectives in using the offset and POD-related services.
- ***SSA staff.*** We will conduct interviews with SSA staff who oversee the activities associated with administering the POD benefit offset.
- ***Mathematica survey staff (recruitment).*** We will conduct interviews with Mathematica staff who are processing completed enrollment packets and responding to telephone calls from prospective enrollees.

We will collect four rounds of qualitative data that will provide updates on the progress of POD implementation and service delivery and will feed into the primary reports that we will deliver to SSA (see Chapter X). Round 1 of data collection (early 2018) will focus on understanding the program environment surrounding each VR agency/WIPA provider, including the local service setting, outreach and recruitment efforts, the POD program infrastructure, and early successes and challenges encountered during the pilot period of recruitment. Round 2 (late 2018) will focus on changes in the POD program infrastructure and early service delivery, including work incentives counseling, monthly reporting of earnings and Impairment-Related Work Expenses, and processing of the POD offset. Round 3 (fall 2019) will focus on the status of the POD infrastructure and service delivery during a steady state of operations as well as the results of adjustments made to correct issues identified during the start-up phase of the demonstration. Round 4 (winter 2020–2021) will gather information about service delivery at the end of the demonstration, including successes and challenges encountered, that can help the evaluation team interpret demonstration findings in the final impact evaluation report.

In the remainder of this section, we summarize our planned interviews with each of the four entities above, including our methodological approach to collecting qualitative data during each round. Our general approach balances the need to collect comprehensive information to support each of our analyses with practical considerations to collect the data in a cost-effective fashion that minimizes any disruption of program delivery and the interview burden on each of the four

groups above. In subsequent deliverables to SSA, we will present data collection protocols that we will use for the first round of data collection, including the specific questions that we will ask each of the four stakeholder groups above and probes to elicit more nuanced information from respondents.[13] We conclude with a summary of the four rounds of data collection and describe the implications about how each round supports the analyses described in Chapters VI through IX.

A. Approach to data collection

Mathematica and Insight Policy Research will lead the qualitative data collection activities in each of the states, with a designated state liaison guiding the effort for each state. The state liaison will participate in monthly meetings with demonstration partners (state VR agency/WIPA managers, SSA management, Virginia Commonwealth University technical assistance liaisons, and Abt Associates) over a 36-month period to stay abreast of any changes in the implementation of POD and the program environment, share updates about Mathematica's evaluation and data needs, and gather contextual information the liaisons will systematically track in the POD SharePoint site. We will gather information in these meetings to inform the content of our interviews with state VR/WIPA managers who will oversee state VR/WIPA functions and also the POD counselors.

As will be discussed in more detail below, the state liaisons will take several steps to maintain high quality data collection. Before conducting interviews, we will train all interviewers on the interview protocols, review the research questions and data points we will cover during each interview, develop data collection procedures, and review interviewing best practices. We will pilot all training beforehand so that interviewers ultimately receive consistent training. Interviewers will record each interview so that all interviews will be transcribed verbatim for analysis and so that we can provide feedback to interviewers to facilitate consistent recording of information. After each round of data collection, the task lead will do a high-level review of transcribed notes for missing or inconsistent data. All data (including recording and transcript notes) will be stored on a secure server, accessible only to relevant project staff.

1. Program document review

As a starting point for each round of data collection, we will review existing program documents and training materials. The program documents will include unpublished materials that are directly relevant to the intervention design, site reports, and plans for technical assistance, along with other implementation materials (for example, procedural manuals,

[13] In a separate deliverable, we summarize our data collection plans for the qualitative analysis (Deliverable 6.3–Data Collection Plan), including discussion topics for each of the four groups of respondents as well as our planned coding schemes that we will use to organize and analyze the qualitative data collected during each round (summarized in Chapter VI). Additionally, we have a separate deliverable (Deliverable 6.4–Site Visit Template) that presents interview protocols for the first round of site visits in early 2018, an observational checklist that will be used to observe site operations in the POD processing center and POD call center, and site visit agendas we are working to develop with each demonstration site prior to recruitment.

screening and assessment tools, recruitment brochures, and formal meeting notes). [14] Before each round of data collection, we will review these documents to solidify our understanding of POD implementation procedures and salient characteristics of the implementing entities in each POD state. This review will help us to identify gaps in our knowledge about each state's demonstration projects, allowing us to identify questions to be included during site visits and telephone interviews with key informants described below. We will also use these documents to tailor protocols to reflect how the demonstration is being implemented in each state, including which functions are performed by state VR agency/WIPA provider staff versus by a centralized support unit at Abt.

2. Balancing site visits and telephone interviews

During every round of data collection, we will conduct phone interviews with key stakeholders and, in select rounds, we will also hold in-person site visit interviews. One reason for the mixed methods approach to qualitative data is to reflect the structure of the organizations delivering services. Specifically, the POD implementation providers include a mix of WIPA and VR service providers who are geographically dispersed and, in some cases, do not include a physical location to provide services. For example, some POD implementation providers have a physical location where POD subjects can obtain work incentives counseling in person, whereas other POD implementation service providers deliver work incentives counseling through distance-based approaches via telephone or videoconferencing.

We will tailor our approach to using phone and in-person visits based on the size of the catchment area in each state and whether the key respondents are geographically dispersed. We will conduct in-person site visits when respondents are located in relatively close proximity (within 90 miles) of one another, and we will conduct telephone interviews when respondents are geographically dispersed (since in-person data collection in these circumstances would be costly and impractical). As of the writing of this report, we expect to conduct in-person site visits to six states where key respondents are situated in close proximity to where the implementation provider is physically located.[15] In the remaining two states (Nebraska and Texas), we will explore with the VR agency/WIPA managers during the planning call whether conducting in-person site visits will be feasible. If respondents are geographically dispersed, we will use telephone interviews to obtain information about the program environment and service delivery. The dedicated state liaison for each site will conduct the site visits/interviews and will maintain all contact with VR agency/WIPA provider points of contact in their respective designated states.

When a physical location is accessible, we plan to conduct in-person interviews at key points in the demonstration—in the first round three months after recruitment begins and the third round, one year after enrolling the last participant. Specifically, the first round of in-person interviews is necessary to establish a working relationship with implementation staff and observe site operations during enrollment. We will do another set of in-person interviews during Round 3

[14]We will include information from three unpublished reports from the Abt implementation team about the intervention design, site reports, and technical assistance (Abt Associates 2017a, Abt Associates 2017b, and Abt Associates 2017c),

[15] These assumptions are subject to change based on new information learned from VR agency/WIPA managers during the planning call that will be held in January/February 2018.

to observe service delivery during a steady state, which will allow us to assess the fidelity of service delivery. We plan to conduct telephone interviews in the interim years between in-person site visits as a targeted, efficient means of collecting data. The state liaisons will facilitate the discussions by phone because staff will have preexisting relationships with the key respondents.

3. Semi-structured interviews with POD treatment subjects

As a supplement to our data collection in each round, we will conduct semi-structured interviews with POD treatment subjects in Rounds 2 and 3 to gain their perspective about the POD offset and service delivery. These interviews will allow our team to recruit subjects based on specific criteria other than geography, to collect richer contextual details, and to elicit more in-depth information than in a group setting.

Ahead of the interviews, we will identify a sample of potential interviewees who have characteristics or experiences about which we are particularly interested in learning. For example, we plan to stratify the sample to reach individuals who formally withdrew from the demonstration to learn about their reasons for exiting the demonstration. We will also attempt to contact a small number of SSDI beneficiaries from the recruitment pool who received a recruitment packet but declined to enroll in POD, to understand their reasons for not wanting to volunteer for the demonstration.[16] To learn about reasons behind work behavior, we plan to target offset users, along with subjects whose level of work was sufficient to reach termination of benefits. For each round of interviews, we will complete nine interviews in each site, totaling 72 interviews per round. We will select the subjects from our management information system (described in Chapter IV). After each interview, we will mail interviewees a $40 incentive payment in appreciation of their time and participation, and will inform them of this in advance of their participation.

4. Preparation and reporting activities

Our qualitative data collection for all rounds requires detailed planning and effective coordination with demonstration partners in each of the POD states. Approximately three months before the first round of site visits in early 2018, the state liaisons will participate in a conference call with the Virginia Commonwealth University site director and VR agency/WIPA manager in each POD state to discuss Mathematica's data collection plans. Shortly after the call, the designated state liaison will send an email to the state VR agency/WIPA provider point of contact for each POD state. The email will describe whether the activities will occur via phone or in person, identify the approximate time frame for a visit, and request a date for a planning meeting via telephone to discuss the logistics of the site visit and all site visit activities. During the planning meeting with the state VR agency/WIPA provider point of contact, we will discuss the schedule for the interviews (for example, length of interviews with each key informant and each informant's role and responsibilities within the organizational structure of the state VR agency/WIPA provider) and learn where each key informant is located. After these initial meetings, the state liaisons will follow up by email and telephone to coordinate logistics. We will repeat this planning for each meeting, though the site visits will have a more intensive planning

[16] It might be difficult to contact these beneficiaries given their lack of contact with the demonstration. This could limit the sample we include who respond to these questions. We will document our outreach to this group to understand if they are more difficult to contact relative to other groups.

effort given the more in-depth natures of these visits (see Appendix C for more details about planning for site visits).

We will code the final interview transcripts and observation notes from all site visits and informant interviews using standard coding schemes, based on the evaluation research questions of interest, key components of the program, and the analytic framework described in Chapter VI. For the analysis, we will search, retrieve, and sort information by different code combinations and create text files to identify themes and cross-cutting findings.

B. Priority topics for each stakeholder group

In Exhibit V.1, we summarize our approach to data collection using the methods described above (program documents, site visits, telephone interviews, and semi-structured interviews) for our four stakeholder groups. As shown in the table, each round of data collection has a specific purpose that will feed into each report. Below, we provide additional details on the topics we will cover with our four stakeholder groups for each round of data collection.

Exhibit V.1. Process data collection activities

Process data collection round/key content	Implementation management and site staff[a]		POD treatment subjects	SSA staff	Mathematica survey staff	All
	Site visits	Telephone interviews	Semi-structured interviews	Telephone interviews	Telephone interviews	Review of program documents and observations
Round 1 (winter 2018)	X	X		X	X	X
	Key content: Program environment, site background, outreach efforts, development of the POD infrastructure, recruitment and enrollment during the pilot phase of operations, and early successes and challenges					
Round 2 (fall 2018)		X	X	X	X	X
	Key content: POD recruitment and enrollment, progress developing the POD infrastructure, subjects' perspectives on POD, adjustments made to correct issues identified in the Early Assessment Report					
Round 3 (fall 2019)	X	X	X	X	X	X
	Key content: The POD infrastructure, staff use of the MIS, work incentives counseling, subjects' perspectives on POD, monthly reporting of earnings and Impairment-Related Work Expenses, processing of the offset, and successes and challenges					
Round 4 (winter 2020–2021)		X		X	X	X
	Key content: The POD infrastructure, staff use of the MIS, work incentives counseling, monthly reporting of earnings and Impairment-Related Work Expenses, processing of the offset, off boarding from POD, successes and challenges, and lessons learned					

[a] Implementation management and site staff include POD counselors delivering work incentives counseling and other supports to treatment subjects, VR agency/WIPA managers supervising the POD counselors, local VR agency staff, the state VR director who signed a letter of intent to initially participate in POD, and local stakeholders with knowledge of the employment service system in the locale.

1. Implementation management and site staff

We will conduct interviews with implementation management and site staff in each round, including Abt staff housed in the indirect support units (POD processing center and POD earnings support), Virginia Commonwealth University staff who are providing training and technical assistance to POD counselors, and Abt implementation management staff who oversee operations.

Our first two rounds of data collection will include information at the start and end of recruitment, respectively (Exhibit V.2). At the start of recruitment (Round 1), we will collect in-person and telephone information during the recruitment pilots about the program environment; perceptions of the recruitment outreach; and implementation information about their infrastructure, such as staffing. We will also talk to these staff to learn about the service environment and other programs that could affect how POD subjects access the offset or related services. We will report findings to SSA and share information with the implementation contractor staff that might facilitate improved service delivery. We will start our second round of data collection (Round 2) toward the end of the recruitment period. At this stage, we expect the implementation to have some early lessons from service delivery in serving a larger sample than was available in Round 1 given our visits coincide with the completion of the full rollout of the recruitment period. We will also assess changes in the program environment that might affect planned service delivery and early start-up activities, including enrollment.

Exhibit V.2. Summary of discussion topics with implementation management and site staff, by data collection round

Discussion topics	Data collection round			
	1	2	3	4
Program environment				
General: Employment environment, state policies, and other state-specific contextual features	X			
Changes: Changes in program environment during the demonstration		X	X	X
Recruitment, enrollment, and random assignment				
Perceptions of direct and indirect outreach strategies	X			
Implementation of the intervention and service delivery				
Planning and early implementation	X	X		
Infrastructure (such as staffing, training, and data systems)	X	X	X	X
Training and provision of technical assistance	X			
Coordination with SSA, Abt, and state agencies	X	X	X	X
Administration of the POD benefit offset and WIC service delivery	X	X	X	X
Assessment of implementation fidelity (that is, interaction with subjects, processing of earnings and Impairment-Related Work Expense information, interaction with SSA)		X	X	X
Barriers and facilitators		X	X	X
Costs of implementing intervention components and other supports				X

Our final two rounds of data collection will focus on service delivery at the middle and end points of the demonstration. Approximately midway through service operation (Round 3), we will conduct a round of site visits and interviews to assess intervention services during full implementation. During these interviews, we will update the information we obtain about the

general program environment and POD infrastructure during the first two rounds. We will also conduct in-depth interviews with implementation staff to assess the fidelity of service delivery. One important example of our interviews will include the POD processing center and POD call center (housed in Abt Associates' office located in McAllen, TX) to observe the operations of these centralized POD support teams that help process monthly earnings and Impairment-Related Work Expense information for POD offset users. These centers play key roles in the program: the POD processing center will receive, process, and update all earnings and Impairment-Related Work Expense documentation that treatment subjects submit monthly and identify beneficiaries that require follow-up by either the call center staff or an assigned counselor because the earnings or Impairment-Related Work Expense documentation did not meet quality assurance standards; the POD call center will respond to calls from treatment subjects, SSA, project partners, and the general public. The POD call center will also make telephone calls to treatment subjects with earnings over the POD threshold to remind them to report monthly earnings information. We will also interview telecommuting POD earnings support staff who will review a subset of earnings records for POD offset users for quality assurance before they are submitted to SSA. Finally, toward the end of service delivery, we will use our telephone interview to update information about service delivery from the third round and collect new information about closeout. We will also collect cost information from implementation staff that is not included in Abt's Implementation Data System, such as management costs for oversight, for use in the benefit-cost analyses (detailed in Chapter IX).

2. POD treatment subjects

To gain the perspective of program participants, we will hold two rounds of one-on-one semi-structured telephone interviews with POD treatment subjects that coincide with the second and third rounds of data collection (Exhibit V.3). We will use an interview protocol designed to yield information about beneficiaries' decisions related to POD participation and work. We will design the interviews to complement the information in the follow-up surveys (described in Chapter IV). During the first interview, the protocol will include questions on beneficiaries' motivation for volunteering for POD; their perspectives on the outreach, recruitment, and enrollment processes; initial contact with demonstration staff; perspectives about the work incentive counseling services they received; and their understanding of the POD offset rules. For those beneficiaries who declined to enroll in POD, the questions will uncover their reasons behind not wanting to volunteer in the demonstration. The second round of interviews will capture participants' perspectives of their participation in POD including work incentives counseling, earnings reporting, and benefit adjustment; their attitudes toward employment and work experiences; and POD areas in need of improvement.

Exhibit V.3. Summary of discussion topics with treatment subjects, by data collection round

Discussion topics	Data collection round			
	1	2	3	4
Recruitment, enrollment, and random assignment				
Beneficiaries motivations for enrolling in POD		X		
Perceptions of direct and indirect outreach strategies		X		
Implementation of the intervention and service delivery				
Perceptions of work incentives counseling		X		
Understandings of new rules/POD offset		X	X	
Motivations for using or not using POD offset		X	X	
Areas for implementation improvement			X	

3. SSA

We will conduct telephone interviews with SSA Processing Center staff who administer the POD offset to gain their perspective on the offset administrative processes during each interview round (Exhibit V.4). We will obtain information from Processing Center staff about whether the offset is being administered as envisioned. Questions include whether processes are in place for reporting earnings and Impairment-Related Work Expense information to SSA, whether the systems support timely and accurate adjustment of benefits under POD, and any issues around the automated reconciliation process that is run annually to establish the exact offset amount that should have been applied to each treatment subject's SSDI benefits during the prior year.

We will also speak with SSA management staff who oversee implementation and evaluation of the demonstration. We expect these staff will provide input on the administration of the POD offset and POD work incentives counseling services, facilitators to implementation, and challenges encountered that might have influenced implementation of the intervention in the demonstration sites. During Round 4, we will also collect cost information from management staff that we can include in our benefit cost analyses.

Exhibit V.4. Summary of discussion topics with SSA, by data collection round

Discussion topics	Data collection round			
	1	2	3	4
Implementation of the intervention and service delivery				
Planning and early implementation	X	X		
SSA infrastructure (for example, staffing and data systems)	X	X	X	X
Processes to reporting earnings and Impairment-Related Work Expense information to SSA; systems to support timely benefit offset adjustments; automated reconciliation		X	X	X
Other costs to SSA (such as administration)				X

4. Mathematica staff

We plan to speak with Mathematica survey staff who are conducting recruitment and enrollment operations in the first two rounds (Exhibit V.5). These staff will provide information on processes related to obtaining enrollment packets and any issues in processing eligibility files,

and obtain qualitative feedback on the value of the POD toll-free number during the enrollment period. Finally, we will talk with survey staff who processed the consent form to gain an understanding of the general processing issues and wait times associated with receiving mailings and contacting POD subjects.

Exhibit V.5. Summary of discussion topics with recruitment and enrollment staff, by data collection round

Discussion topics	Data collection round			
	1	2	3	4
Recruitment, enrollment, and random assignment				
Survey Operations Center infrastructure (for example, staffing, training, and data systems)	X	X		
Beneficiary calls to toll-free number	X	X		
Receipt of mailings, processing of consent forms, and random assignment	X	X		

C. Evaluation implications

In Exhibit V.6, we summarize each of the data sources described above by its planned use in each of the analyses in subsequent sections and updates by data collection round. As shown in the exhibit, the data collection includes four key stakeholder groups (implementation management and site staff, POD treatment subjects, SSA staff, and Mathematica staff). We plan to include perspectives of implementation management and site staff and SSA in all rounds and information from POD treatment subjects and Mathematica recruitment staff in select rounds. Exhibit V.6 and the corresponding summary of quantitative data sources in Exhibit IV.4 provide a full summary of all the planned qualitative and quantitative data sources for the evaluation, respectively. In the subsequent chapters, we provide more details about how we plan to integrate these data into each of our analyses.

Exhibit V.6. Summary of qualitative data for evaluation

	Process analysis	Participation analysis	Impact analysis	Benefit-cost analysis	Data updates (data collection rounds)
Implementation management and site staff	X	X		X	All rounds (1, 2, 3, and 4)
POD treatment subjects		X			Rounds 2 and 3
SSA staff	X	X		X	All rounds
Mathematica staff		X			Rounds 1 and 2

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

VI. PROCESS ANALYSIS

This chapter describes the POD process analysis, which we have designed with the broad objective of understanding the implementation of the POD benefit offset and its associated services, the POD enrollment process, and the context in which POD operates. More specifically, the process analysis will help us answer research questions in the following domains:[17]

- ***Program environment.*** We will present information about the program environment, which will provide context for our process findings related to service delivery, as well as set the stage for other analyses. We will summarize general employment environment, state policies, and other state-specific contextual features during the time of the demonstration that might influence eventual outcomes.
- ***Recruitment, enrollment, and random assignment.*** We will discuss the approach to recruitment, including whether there were any important deviations in the process of recruitment, enrollment and random assignment from the initial design presented in Hock et al. (2017). To obtain this information, we will summarize perspectives from staff who supported recruitment at Mathematica and obtain the perspectives of POD subjects to understand their motivations to enroll in the demonstration. We will also attempt to contact a small number of SSDI beneficiaries from the recruitment pool who received a recruitment packet but declined to enroll in POD, to understand their reasons for not wanting to volunteer for the demonstration.
- ***Describe and assess implementation of the intervention.*** The largest component of our process analysis will focus on implementation and service delivery. We will document the background (such as years of relevant experience and training certifications) of VR/WIPA managers and also the POD counselors, paying close attention to potential differences between sites. We will then summarize process related findings from the perspective of program implementation managers and staff, SSA, and beneficiaries about the aspects of program implementation that were most successful and then review whether the implementation team and SSA delivered services and made offset adjustments with fidelity to the original design.

There is substantive overlap between the domains here, in the participation analyses, and in the impact analyses. For example, if the participation analysis finds only a small percentage of treatment subjects using the POD benefit offset in a given state, the process analysis could provide insight into barriers the states encountered in processing and reporting monthly earnings and Impairment-Related Work Expenses. Likewise, identifying internal or external factors that might have influenced implementation of the POD benefit offset or work incentives counseling will enable the impact team to assess intervention components that might relate to the eventual outcomes. Insights into the specific facilitators to successful implementation and barriers that hindered implementation of the POD benefit offset can inform the impact team's thinking about why no effects were found (if applicable) or potential sensitivity analyses that might be useful to conduct.

[17] The full universe of research questions that will be addressed by the process analysis is available in Appendix D of the data collection plan (Deliverable 6.3).

The process analysis' role in the early assessment will also provide formative feedback that SSA could choose to share with the states and implementation team. We will provide a detailed overview of the program environment in the Early Assessment Report to identify any potential changes that might improve the implementation of the POD work incentives counseling and the benefit offset. In the subsequent interim and final reports, we will review whether the program environment experienced any changes that acted as mitigating factors in influencing program impacts.

A. Research questions and data sources

To provide a comprehensive assessment of each domain, we will draw on information from each of the data categories summarized in Chapters IV and V (Exhibit VI.1). We will use quantitative data sources for indicators of adherence to plans and fidelity to the original design as well as measurable differences between states. We will use qualitative data to understand why and in what ways implementation deviated from the original plans and the barriers and facilitators to implementation.

Exhibit VI.1. Process analysis research questions and data sources

Question	Data source			
	Program data	Management information system (MIS) data	Survey data	Qualitative data
Describe and assess program environment				
What were the general employment environment, state policies, and other state-specific contextual features (such as public transit) during the time of the demonstration?				X
How prevalent was engagement in the Ticket to Work program, and how were employment networks structured? What existing employment services are available to SSDI beneficiaries before POD?	X			X
How has the environment changed with program implementation?				X
How did state VR agencies and WIPA providers previously serve SSDI beneficiaries, and how did this change during the demonstration period?				X
Describe the recruitment, enrollment, and random assignment processes				
What percentage of mailers were returned from the direct outreach strategy?			X	
How was random assignment implemented? What strategies were used to maintain the integrity of random assignment?		X		X
What strategies were used to inform SSDI beneficiaries about POD? How did outreach and enrollment differ for concurrent beneficiaries (those who receive SSI and SSDI)?		X		X

EXHIBIT VI.1. *(continued)*

Question	Data source			
	Program data	Management information system (MIS) data	Survey data	Qualitative data
Have the targeted or broad outreach strategies been effective in reaching prospective enrollees? What aspects of the recruitment strategy were successful/not successful?		X		X
Did prospective enrollees complete the baseline surveys properly?		X		
What efforts were made to inform local stakeholders (VR staff, SSA field office staff, WIPA's, etc.) and service providers about POD?		X		X
How was informed consent implemented with prospective enrollees?				X
How do subjects view POD? What motivated subjects to enroll in POD?			X	X
What factors affected volunteer rates?	X	X		X
Describe and assess implementation of the intervention				
What are the contextual factors that are facilitating or hindering state VR agencies' and WIPA providers' ability to make implementation changes during the demonstration period?		X		X
What aspects of program implementation have been successful and why? What facilitated timely reporting of earnings and Impairment-Related Work Expenses?	X	X	X	X
How are work incentives counseling services delivered to treatment subjects?				X
How do POD counselors support earnings reporting for employed treatment subjects?				X
What types of work supports and services are delivered to treatment subjects?		X		X
What facilitated benefit offset use?	X	X	X	X
What are the administrative (that is, implementation, operations, or systems) challenges in providing the intervention?				X
How quickly are benefit payments adjusted for offset use, and what role did state VR agencies and WIPA providers play in making these adjustments?	X	X		X
What are the major reasons for benefit adjustment delays, if any?		X		X
What is the incidence and frequency of overpayments for benefit offset use relative to the control group?	X	X		

Notes: See Chapter IV for the specific data sources included in the program, MIS, and survey data. See Chapter V for the specific data sources included in the qualitative data. See Appendix D in Deliverable 6.3 for the full universe of research questions that will be addressed in the process analysis.

B. Analytic approach

A potential challenge for the process analysis is to synthesize information from several diverse perspectives in a way that minimizes subjectivity. We will address this issue by integrating quantitative data and using a framework based on the implementation science literature that will provide a consistent structure to guide qualitative data collection and analysis (Alexander and Hearld 2012). Specifically, we plan to use the Consolidated Framework for Implementation Research (CFIR) to systematically identify factors that might influence implementation effectiveness. The CFIR is a conceptual framework that guides systematic analyses of multilevel and diverse contexts in which interventions are implemented and helps explain the myriad factors that might influence intervention implementation and effectiveness (Damschroder et al. 2009). We will also integrate important qualitative perspectives that emerge from our data collection. These other qualitative perspectives might emerge because our data collection team identifies a new theme that was not anticipated in the development of the protocols. Below, we describe our plans to integrate all sources of quantitative and qualitative data into our analyses.

1. Quantitative Data

As a starting point, we will integrate quantitative data into the process analysis to inform the development of the data collection protocols. For example, tabulations of the number of enrolled study subjects for each state and the number of T1 and T2 subjects using the benefit offset and engaging in counseling for each state are important indicators of the original assumptions about recruitment and fidelity to the intervention's design. The combination of quantitative and qualitative data could be especially important for cross-cutting analyses in connecting findings from the process, participation, and impact findings. For example, in the BOND evaluation, the evaluation team noted that beneficiaries had difficulties understanding the BOND work rules, which might have limited offset usage and ultimately influenced impacts (Wittenburg et al. 2015).

2. Framework for studying implementation

We will use the CFIR to guide the coding and analysis of the qualitative data collected during interviews with VR agency and WIPA managers, POD counselors, and demonstration partners, and to report our findings related to assessing POD implementation. By specifying a set of factors demonstrated in prior research to influence effective implementation, the CFIR will increase the relevance of our research findings for informing implementation.We will report findings from our analysis of factors that might influence implementation in the Early Assessment and Random Assignment Reports to provide SSA Processing Center staff, state VR/WIPA provider staff, POD counselors, and POD implementation staff with formative feedback on program operations.

We will develop codebooks to reduce and organize the qualitative data collected during interviews. We provide codebook examples in Appendix D. We will use the codebooks to guide coders through three decisions for each data segment, each of which includes an interview question and response. The coder will (1) use the research question codebook to determine which of the research questions is relevant to the information being discussed; (2) use the program components codebook to determine which of the components are being discussed; and

(3) identify which one of the five CFIR domains is reflecting the principal implementation theme in the data (for example, intervention characteristic).

In the initial stages of coding, a team of three to four coders will review interview transcripts and code data together. During this process, the team will refine code definitions, develop coding rules, and resolve disagreements to achieve consistency in their application of the codes to the data. After achieving consistency in applying codes to the data, we will divide the remaining transcripts among the team. Throughout the coding process, senior evaluation team members will review a subset of coded interviews for accuracy so that coders are consistently applying codes to topics that emerge from the data. We will use NVivo to code and organize the data for analysis.

3. Other qualitative reporting

While we will structure our qualitative protocols to facilitate coding using CFIR, we anticipate that our four stakeholder groups could raise issues not fully captured in our protocols that could be especially informative to the process analysis. As one simple example, we anticipate the qualitative perspectives from implementation staff on lessons learned could be particularly valuable in summarizing process findings to SSA about ways to improve the implementation of POD for, say, a national program. For these reasons, we will report other qualitative perspectives from our key stakeholders, particularly to provide perspectives about lessons from POD. When we report other qualitative findings, we will note the perspective in our reporting to SSA. For example, in BOND, we found that some implementation site staff reported concerns that beneficiaries were uncertain about the legitimacy of the intervention, which was reported on a few sites. The qualitative team did not have a question about the legitimacy of BOND in their protocols but still reported this important qualitative finding in the reports with the necessary caveats as it provided some insight into the beneficiary perception of the demonstration (Derr et al. 2015).

C. Reporting findings

The process analysis hinges on our ability to collect information systematically on a broad range of topics and from a large number of sources and to organize the information within a framework that addresses the three general process questions outlined in Exhibit V.1. Below, we highlight our approach to addressing these questions using the analytic methods noted above.

1. Program environment

The program environment refers to the features of the employment environment and other contextual features of states. The process analysis will assess the extent to which the program environment appears to affect the use of the POD benefit offset or the operation of other state or federal entities and changes during the demonstration period. Prior to and during the first round of site visits, we will describe and document the program environment in each state to understand: (1) which agencies serve SSDI beneficiaries, (2) the types of employment services and supports that are available to treatment subjects before and during the demonstration period, and (3) treatment subjects' past use of workforce incentives and the services offered through state VR agencies and other Employment Networks. For our early reports to SSA, we will

document the local economic and employment service environment, which might, in turn, affect beneficiaries' opportunities to earn wages over the POD threshold.

In Exhibit VI.2 we present a summary table that illustrates our structured approach to summarizing the program context between the eight POD sites. The table compares POD counseling providers, economic indicators and the service environment, which allows readers to obtain an at a glance comparison of each site's provider characteristics with the economic and service environment. In subsequent reports, we will then provide updates about changes in the program environment (see Chapter X for a full summary of reports).

2. Recruitment and enrollment processes

To understand and document the recruitment and enrollment processes, we will interview Mathematica staff tasked with processing completed enrollment packets (including informed consent) and baseline questionnaires that prospective enrollees complete before random assignment. We will also speak with Mathematica staff who respond to calls placed to the POD toll-free number during the enrollment period.

We will also capture treatment subjects' perspectives from the semi-structured interviews at the end of round 2 of data collection (see Chapter V). We will use the interviews to obtain information about their initial impressions of outreach and recruitment strategies, their motivations for enrolling in POD, and their experience with the enrollment processes. We will also explore the related question of what were their motivations for subsequently using or not using the offset, such as whether financial incentives or simplified rules were more salient for their decision.

In Exhibit VI.3, we present a summary of possible measures capturing information about outreach, recruitment, and enrollment activities during the enrollment period for a given month (e.g., January 2018). We plan to analyze the recruitment and enrollment metrics at two points in time. In March 2018, we will analyze data on the recruitment and enrollment metrics from January 2, the start of enrollment, through February, 2018, the end of the pilot recruitment period. In February 2019, we will analyze data on the same set of recruitment and enrollment metrics representing March 1, 2018, the start of full rollout, through December 31, 2018, the end of the full enrollment. We will use this exhibit to quantify the level of effort and success rate of each stage of study intake, and document differences across states. We will also document any changes to the recruitment and enrollment processes that we observe during our first and second site visits in the Recruitment and Random Assignment Analysis Report due in 2019.

Exhibit VI.2. POD site-level contextual characteristics

Site	POD Counseling Providers				Economic Indicators				Service Environment			
	# of FTEs	Case assignment method	Implementing entity (VR/WIPA/ lower tier subcontractor)	Service coverage in site (full state/ partial)	Unemployment rate (Dec YYYY)	Change in unemployment rate since Dec YYYY	Employment rate among people with disabilities in	Change in employment rate among people with disabilities in YYYY	Operating under order of selection	Reported delays in accessing services	Reported length of delays	Percent of beneficiaries who are Ticket to Work users in YYYY
AL												
CA												
CT												
MD												
MI												
NE												
TX												
VT												

Exhibit VI.3. Indicators of implementation outreach, recruitment, and enrollment, as of February 2018

	AL	CA	CT	MD	MI	NE	TX	VT	All states combined
Indirect Outreach Strategies									
Number of webinar attendees, by participant category (EN staff, VR staff, WIPA, POD Contractor, SSA, other)	---	---	---	---	---	---	---	---	
Number of hits to POD website during enrollment period	---	---	---	---	---	---	---	---	
Direct Outreach Strategies									
Number of prospective enrollees									
Percent prospective enrollees successfully reached by initial mailing									
Percent prospective enrollees reached by initial mailing who subsequently enrolled									
Percent prospective enrollees successfully reached by follow-up outreach (phone or mail)									
Percent prospective enrollees successfully reached by follow-up outreach who subsequently enrolled									
Percent prospective enrollees not reached by follow-up outreach attempt (i.e., bad telephone number)									
Enrollment									
Enrollment target									
Percent enrollment packets returned									
Percent enrollment packets complete									
Percent enrollment target reached									

Note: The measures are subject to change as design and data collection plans are refined further.

3. Implementation of the intervention and fidelity to the service model

Barriers and facilitators to implementation. The CFIR framework includes the following five domains, each of which may affect an intervention's implementation:

1. **Intervention characteristics**, which are the features of an intervention that might influence implementation.
2. **Internal context and setting of the implementing organization**, which includes features of the implementing organization that might influence implementation.
3. **External environment and context**, which includes the features of the external context or environment that might influence implementation.
4. **Characteristics of individuals involved in implementation** that might influence implementation (such as education of staff, years of relevant experience, training and certifications, percent of time spent working on POD).
5. **Implementation processes**, which includes strategies or tactics that might influence implementation and mid-course changes to implementation.

Using the CFIR to guide our coding and analysis will help us systematically examine key facilitators and barriers to implementation.

Exhibit VI.4 presents how we will apply this framework to POD to identify facilitators and barriers to implementing each program component (denoted by "F" and "H" in the table) for a hypothetical example. After organizing the data into analytic matrices for each program component and CFIR domain combination, we will examine patterns of facilitators and barriers to implementing the different program components as they emerge across the eight VR state agencies and WIPA providers serving as POD sites; these matrices will serve as the foundation of our analyses featured in future evaluation reports for SSA.

To illustrate how we will intrepet findings from this exhibit, take the examples under the characteristics of POD domain. In this example, one finding the process team would report is that development of benefits summaries and work incentive plans are key faciliatators in the implementation of POD in part because these processes are compatible with the core mission of the VR agencies and WIPA providers providing work incentives counseling under POD. In contrast, another finding is that the complexity of program requirements were cited as barriers to reporting monthly earnings timely and requesting appeals of benefit adjustments.

Assessment of implementation fidelity. Our analysis of the implementation of the POD intervention will also include a thorough assessment of fidelity of implementation to the intended POD program model, drawing on programmatic data provided by Abt, recruitment and enrollment information, and also qualitative data collected from key informants. We will (1) describe the fidelity with which the intervention is being implemented in all eight POD sites, (2) assess the extent to which fidelity in each site is consistent and the services available are uniform within and across sites, and (3) identify areas where implementation could be supported by additional resources.

Exhibit VI.4. Illustrative table presenting facilitators and barriers to implementation of the POD program components

CFIR domain	Components of work incentives counseling							
	Onboarding	Reporting monthly earnings	Requesting appeals/ waivers	Developing benefits summary & analyses	Developing work incentives plans	Providing ongoing work incentives counseling	Supporting diverse and underserved populations	Transitioning subjects out of POD
Characteristics of POD								
Compatibility with state VR agencies' and WIPA providers' objectives/mission				F	F			
Perception of complex or unclear program requirements		B	B					
Internal context and state VR agency setting								
Experience with work incentives counseling				F	F	F		
Lack of leadership support	B					B		
External environment and context								
Partnerships with other VR agencies or providers in the state	F			F			F	
Complexity of subjects' needs						B	B	B
Characteristics and attitudes of POD program staff								
Knowledge of local service environment and state programs (that is, Medicaid, SNAP, housing supplements)				F			F	
Counselors' skepticism about time-limited benefits	B		B			B		
POD implementation process within the state VR agency or WIPA provider								
Dedicated POD implementation meetings	F	F	F	F	F	F	F	F
Knowledge of program requirements unevenly shared across counselors	B					B		B

Note: For each POD component, F indicates facilitators and B indicates barriers, where applicable.

We will use three resources as our benchmark against which to assess implementation quality: (1) the statement of work for the implementation contractor (Abt Associates) which specifies the basic structure SSA intends to implement for POD; (2) Abt's Implementation Design Report which will serve as the blueprint for the implementation of the demonstration; and (3) the POD implementation manual developed by Abt for POD counselors and VR agency/WIPA provider staff to use as a resource. The assessment will consider overall implementation as well as variation in implementation across each state.

We will assess two main aspects of implementation, (1) delivery of work incentives counseling and other participant supports and (2) reporting of monthly earnings and Impairment-Related Work Expenses information to SSA and adjustment of DI benefits. Given the role of VR agencies/WIPA providers in implementing the demonstration, it will be of particular interest to understand how they support collection and reporting of earnings information from treatment subjects and the extent to which they deliver work incentives counseling accurately and consistently to T1 and T2 subjects.

We present a summary of proposed implementation measures related to earnings reporting in Exhibit VI.5. For example, under reporting of monthly earnings information to SSA, we will examine the percent of treatment subjects with earnings known to be over the POD threshold with complete earnings records and also the average processing time for submitted earnings records. This assessment is particularly important in understanding the SSA's processes for making adjustments to benefits based on submitted earnings and Impairment Related Work Expense information. Hence, this table provides an assessment of the timeliness of benefit adjustments. The full universe of implementation measures that we will assess to examine implementation fidelity are presented in Appendix D. Findings from our analyses of these measures will be presented in future deliverables to SSA.

Exhibit VI.5. Indicators of implementation context of reporting monthly earnings to SSA

	Fidelity Measure	AL	CA	CT	MD	MI	NE	TX	VT	All Sites Combined
Earnings Reporting										
Number of earnings reporting reminder letters sent to treatment subjects in July YYYY										
Percent of all treatment subjects known to have earnings over the POD threshold in July YYYY										
Percent of treatment subjects known to have earnings over POD threshold who reported earnings timely for July YYYY [a]	X									
Percent of treatment subjects who reported Impairment-Related Work Expenses in July YYYY that were not approved	X									
Percent of treatment subjects who reported Impairment-Related Work Expenses in July YYYY that were approved	X									
Percent of treatment subjects known to have a higher Impairment-Related Work Expense threshold who reported timely for July YYYY [a]	X									
Reporting mode for reports made in July YYYY										
Percent reported electronically										
Percent reported by mail or fax										
Percent reported in person										
Earnings Record Processing										
Average, min and max processing time for earnings records submitted in July YYYY	X									
Percent of submitted earnings records submitted in July YYYY requiring follow-up	X									
Percent of records reviewed in July YYYY that fail quality control review	X									
Average time to complete records that failed quality control review in July YYYY	X									
Percent of treatment subjects known to be over the POD threshold with complete records submitted to SSA timely for July YYYY	X									

Source: Abt Implementation Data System

[a] Abt is expected to submit earnings records for treatment subjects who are working to SSA by the 6th following the reporting month. Each earnings record captures information about high IRWEs or earnings paid during the reporting month.

VII. PARTICIPATION ANALYSIS

This chapter describes the POD participation analysis, which includes the following four components:

- ***Recruitment.*** What percentage of invited SSDI beneficiaries volunteer for POD? How do the characteristics of volunteers compare to non-volunteers? What are common reasons for volunteering or not volunteering?
- ***Withdrawal.*** How many treatment subjects withdraw from the demonstration and at what time and stages? How do withdrawals vary between the two treatment groups? What are the most prevalent reasons for withdrawal? How do the characteristics of those who withdraw compare to continuing subjects?
- ***Offset and service use.*** How many subjects use associated services, such as benefits counseling? What is the incidence and frequency of benefit offset use? How does benefit offset use change over time? What, if any, obstacles prevent subjects from using the benefit offset? How do the characteristics of those who use the offset and other services compare to those who do not?
- ***Earnings reporting and overpayments.*** What is the incidence and frequency of reporting monthly earnings and Impairment-Related Work Expenses? What obstacles prevent timely reporting of earnings and Impairment-Related Work Expenses to the implementation contractor? Do outreach efforts or reminders affect monthly earnings reporting? How many subjects experience improper payments arising from delays in benefit adjustments?

The recruitment questions are about all beneficiaries who are in the solicitation pool. They are of particular interest because of the potential recruiting challenges outlined in Chapter III. If our recruitment efforts do not hit the anticipated targets, the recruitment analysis will provide an in-depth review of the issues associated with recruiting volunteers into POD for demonstration services. For this question, we will assess who is interested in volunteering for POD and, among volunteers, who completes the full set of consent materials and passes the cognitive survey screening questions that are part of the baseline survey (Hock et al. 2017). This analysis will also inform efforts to project impacts to the national beneficiary population.

The next three components relate to participation by treatment (T1 and T2) subjects. We will assess the number of treatment subjects who withdraw, use the offset and associated services, and, for those who use the offset, any issues in reporting their earnings or with overpayments.

Addressing these components will also allow our team to develop cross-cutting analyses, particularly for the process and impacts analyses, but also for the special topic reports. For example, the participation analysis will tabulate and characterize who enrolls in POD or uses the offset, while the process analysis will explore why people do or do not participate. The timeliness of benefit adjustments is another issue that cuts across both the process analysis and the participation analysis. Similarly, assessing the impacts of POD rules on the number of overpayments could provide insights into whether overpayments affect subjects' understanding and subsequent use of the offset in ways that influence ultimate impacts.

A. Research questions and data sources

To comprehensively address each of the research questions above, we will draw on information from each of the data categories summarized in Chapters IV and V (Exhibit VII.1). As shown in the exhibit, we will likely need to draw on multiple data sources to fully address each question. In the remaining table shells below, we highlight how we plan to use information from the specific data sources within these four data categories to address each of the participation research questions.

The answers to some questions, such as the use of the offset, will change over time. For example, we expect that the use of the offset will increase over time as treatment subjects adapt to the POD rules. We will produce some of the tables below in multiple reports, to support our understanding of the dynamics of participant behavior.

Exhibit VII.1. Participation analysis research questions and data sources

Question	Data source			
	Program data	MIS data	Survey data	Qualitative data
Recruitment				
What percentage of SSDI beneficiaries volunteer for POD?	X	X		
How long does it take study subjects to volunteer following the initial recruitment mailing?	X	X		
How do the characteristics of volunteers compare to non-volunteers?	X	X		
What are common reasons for volunteering or not volunteering?				X
Among non-volunteers, how many people were contacted, attempted to contact, and not attempted to contact?		X		
Withdrawal from treatment				
How many treatment subjects withdraw from the demonstration and at what time and stages?		X		
How do withdrawals vary between the two treatment groups?		X		
What are the most prevalent reasons for withdrawal?				X
How do the characteristics of those who withdraw compare to continuing subjects?	X	X	X	
Use of the benefit offset and associated services				
To what extent do subjects receive associated services, such as benefits counseling?	X	X	X	X
What is the incidence and frequency of benefit offset use?		X		
How does benefit offset use change over time?		X		
What, if any, obstacles prevent benefit offset use?			X	X

EXHIBIT VII.1 *(continued)*

Question	Data source			
	Program data	MIS data	Survey data	Qualitative data
Earnings reporting				
What is the incidence and frequency of reporting monthly earnings and Impairment-Related Work Expenses?		X		
What obstacles prevent timely reporting of earnings and Impairment-Related Work Expenses?				X
Do outreach efforts or reminders affect monthly earnings reporting?		X		X
To what extent did subjects experience incorrect payments arising from delays in benefit adjustments?		X		

Notes: See Chapter IV for the specific data sources included in the program, management information system, and survey data. See Chapter V for the specific data sources included in the qualitative data.

B. Analytic approach and outcomes

Below, we document our approach to examining each research question. We focus primarily on descriptive tabulations, but some analyses will use regression models to assess how a participation measure varies by beneficiary characteristics, using the same framework as the impact analysis (Chapter VIII). Additionally, we will use qualitative analyses to provide context on certain outcomes noted in Exhibit VII.1 (for example, reasons for non-volunteering).

1. Recruitment and enrollment

To begin the recruitment analysis, we will present statistics about the number of eligible beneficiaries, recruitment targets, and achieved volunteer rates, both in total and by state (Exhibit VII.2). The information in Exhibit VII.2 should reveal recruitment progress during intake and document the distribution of study subjects across the states once recruitment is complete.

Exhibit VII.2. POD volunteer and subject characteristics, by state

	Volunteer target	Number of eligible beneficiaries	Number of volunteers	Percentage of eligible beneficiaries who volunteer	Percentage of volunteer target achieved
All states					
Alabama					
California					
Connecticut					
Maryland					
Michigan					
Nebraska					
Texas					
Vermont					

Sources: SSA program data and RAPTER.

We will compare the demographic, impairment, and other baseline characteristics of study subjects to non-volunteering beneficiaries in the catchment areas. We will also separately tabulate non-volunteering beneficiaries who were contacted, attempted to contact but not reached, and not attempted. These comparisons will help SSA understand the ways in which POD subjects differ from all SSDI beneficiaries and therefore how lessons from POD may or may not apply to the SSDI beneficiary population. Additional analysis of information from qualitative interviews with POD volunteers will identify themes about why certain SSDI beneficiaries volunteered for POD.

Special recruitment experiment topics

The recruitment experiments will include four variations of participant follow-up that our team designed to stimulate interest in participation, including follow-up phone calls, reminder post cards, mail- back postcards that beneficiaries can use to obtain immediate phone follow-up support, and alternative presentation of benefit scenarios in the recruitment materials (Hock et al. 2017). We plan to test the relative efficacy of each strategy based on a factorial experiment design that maximizes the sample size for each two-way contrast while also allowing us to compare all possible combinations of the strategies.

As outlined in Chapter III and Hock (2017), we will also test recruitment experiments that could produce special topical information for the participation analysis related to recruitment, especially in meeting sample targets. While the time available to summarize the changes is limited in the pilot, we propose a special topical report these tests, particularly if the recruitment experiments are informative to other outreach methods that SSA can use to communicate with beneficiaries for other administrative and operational purposes.

Of particular interest for the recruitment analysis is to assess 1) whether the demonstration can meet its target of 15,000 subjects and 2) understanding who volunteers for the demonstration and why. As outlined in Chapter II, there are strong theoretical expectations that beneficiaries who complete their Grace Period will have more incentive to participate in POD than do other beneficiaries. Our early analysis here will assess if this theoretical expectation holds to form and, if so, the implications for generalizing findings from the evaluation. For example, the findings here could inform whether sufficient sample might be available to construct subgroups for the impact analysis (see Chapter VIII).

Exhibit VII.3 shows the types of baseline characteristics we will calculate for study subjects, non-volunteers, and all eligible beneficiaries in the catchment areas (study subjects and non-volunteers), and the comparisons we will make between the groups. Exhibit VII.3 also illustrates which characteristics we will compare, but we will add other characteristics as data become available and include them in updated versions of this table for later reports. For example, we expect that we can access pre-enrollment earnings and education data to include in the impacts reports that will not be available for the early assessment report.

Exhibit VII.3. Characteristics of study subjects, non-volunteers, and eligible beneficiaries in the catchment areas

Variable	Sample mean			Standardized differences
	Study subjects	Non-volunteers	Eligible beneficiaries	Subjects vs. non-volunteers
Number of beneficiaries				
Gender (percent)				
Male				
Female				
Age (percent)				
20–29 years				
30–39 years				
40–44 years				
45–49 years				
50–54 years				
55–59 years				
Mean age (years)				
Primary impairment (percent)				
Neoplasms				
Mental disorders				
Back or other musculoskeletal				
Nervous system disorders				
Circulatory system disorders				
Genitourinary system disorders				
Injuries				
Respiratory				
Severe visual impairments				
Digestive system				
Other impairments				
Completion of Trial Work Period and Grace Period				
Trial Work Period				
Grace Period				
Years since onset of disability (percent)				
Less than 2 years				
2 to less than 4 years				
4 to less than 6 years				
6 to less than 8 years				
8 to less than 10 years				
10 to less than 12 years				
12 or more years				
Other variables				
Number of auxiliaries in SSDI				
Baseline earnings				
Monthly SSDI benefits ($)				
Number of months received SSDI				
SSDI-only (percentage)				
Concurrent beneficiary (percent)				
F statistic				

Sources: SSA program data and RAPTER.

2. Withdrawing from POD

The analysis of treatment group subjects who withdraw from POD will in some ways mirror the analysis of recruitment. We plan to use similar data sources to understand who withdrew from POD and why.

Exhibit VII.4 provides an example of a withdrawal analysis of POD treatment subjects by treatment status and state. The exhibit informs research questions about whether POD withdrawal rates differed by treatment group and/or region. Similar to the analysis shown in Exhibit VII.3, an exhibit in the analysis regarding withdrawal from treatment will describe the characteristics of those who withdraw from POD and compare them to the characteristics of those who remain. While not shown, if there are a substantial number of withdrawals we will also make comparisons of subjects who withdraw to all POD subjects to assess whether those who withdraw present specific subgroups of beneficiaries.

Exhibit VII.4. Withdrawals from POD as of (date), by state

	T1		T2		T1 + T2	
	Number who withdrew	Percentage of subjects who withdrew	Number who withdrew	Percentage of subjects who withdrew	Number who withdrew	Percentage of subjects who withdrew
All states						
Alabama						
California						
Connecticut						
Maryland						
Michigan						
Nebraska						
Texas						
Vermont						

Sources: SSA program data and RAPTER.

3. Accessing the benefit offset and associated services

Research questions in the third domain investigate the extent to which treatment subjects engage the primary components of the POD intervention—the benefit offset and the associated services. The effects of the intervention derive from whether (as well as the extent to which) treatment subjects use the benefit offset and engage POD's associated services. Therefore, it is important to measure these forms of engagement. Questions regarding associated services, such as the receipt of benefits counseling, will be included among the fidelity measures for the process study. We will use the Abt Implementation Data System, SSA's POD Automated System, and SSA program data to quantify benefit offset utilization.

As one example of offset and service use, we present monthly benefit offset utilization for both treatment groups, both cumulatively and within each month (Exhibit VII.5). We will calculate estimates using the impact analysis framework described in Chapter VIII. Like the other impact estimates, this exhibit will be presented graphically, with detailed estimates based on the framework of this illustrative table in the appendices.

We will also examine characteristics of study subjects who use the benefit offset compared to study subjects who do not, separately for each of the two treatment groups. These comparisons will use a framework analogous to that presented in Exhibit VII.3. As discussed in Chapter VI, the process analysis will further help us understand why individuals use the benefit offset and associated services, as well as possible barriers to use, such as not understanding the rules.

Exhibit VII.5. Utilization of benefit offset and associated POD services

Outcome (percent)	Sample mean		Estimated difference
	T1	T2	T1 vs. T2
Offset use			
Ever used benefit offset	nn.n%	nn.n%	nn.n (nn.n)
Used benefit offset 1 month after enrollment			
Used benefit offset in 2 months after enrollment			
Etc.			
Used benefit offset 1 to 2 months after enrollment			
Used benefit offset 1 to 3 months after enrollment			
Etc.			
Use of associated POD services			
Received benefits counseling	nn.n%	nn.n%	nn.n (nn.n)
Received employment services			

Sources: RAPTER and Abt's Implementation Data System.

4. Earnings reporting and prevalence of overpayments

The final set of questions for the participation analysis examines earnings reporting and related payment measures. Timely and accurate reporting of earnings and Impairment-Related Work Expenses facilitates timely and accurate SSDI benefit payments. If POD treatment subjects frequently encounter delays, inaccurate benefit payments, and later corrections, it may confuse them or undermine their confidence that their benefits will be adjusted correctly when they increase their earnings. In addition, because we expect many treatment subjects to have low household incomes and assets, they may find it difficult to adapt to unanticipated frequent or large changes in their benefit check.

We will use program and process study data to track earnings reporting and improper payments. The process analysis' fidelity study will investigate the incidence and frequency of reporting for monthly earnings and Impairment-Related Work Expenses. Program data will quantify the prevalence and amount of improper payments throughout the life of the study (Exhibit VII.6). For any outreach efforts designed to improve earnings reporting, we will conduct a pre/post analysis looking for changes in associated program measures.

Exhibit VII.6. Prevalence of improper payments

Period	Percentage of benefit offset users with overpayment	Mean overpayment for those with overpayment (in dollars)	Percentage of benefit offset users with underpayment	Mean underpayment for those with underpayment (in dollars)
January–March 2018				
April–June 2018				
July–September 2018				
October–December 2018				
January–March 2019				
April–June 2019				
July–September 2019				

Sources: RAPTER and SSA program data.

VIII. IMPACT ANALYSIS

Our impact analysis will address the following research questions:

- ***What is the impact of the POD benefit design on four primary outcomes?*** We define these outcomes based on the theoretical justifications outlined in Chapter II to include substantive employment (defined as annual earnings above the SGA level), earnings, annual SSDI benefit payments, and total annual income (SSDI benefits plus earnings). We can measure each of these outcomes from the administrative data, which means we can examine these outcomes throughout the entire demonstration period, and SSA can use its administrative data to examine long-term primary outcomes beyond the demonstration.
- ***What is the impact of the benefit offset on secondary outcomes and subgroups?*** We define the secondary outcomes as all of the other potential impacts of the offset on outcomes shown in our discussion of the conceptual foundation of POD in Chapter II. These outcomes cover the following domains: employment, earnings, income (including supports other than SSDI), benefits from SSDI and other services, and other outcomes (such as health and quality of life). We use a combination of survey and administrative data to provide a comprehensive assessment of the impact of the offset within these domains. Additionally, for all primary and secondary outcomes, we assess whether the impact of the offset varies across key subgroups.
- ***What are the differences in impacts between the T1 and T2 groups on these outcomes?*** In addressing these questions, we define impacts as the collective impacts of the new POD benefit offset, associated benefits counseling, and other administrative changes (which we will refer to collectively as "benefit offset impacts"). In the remainder of this chapter, we describe our analytic approach to addressing the aforementioned impacts using administrative and survey data, including our plans to present impacts by key subgroups, our ability to detect impacts for the overall sample and subgroups (statistical power), and our approach to presenting findings and generalizing them for policy purposes.

We will examine these research questions with comparisons between each pairwise combination of the T1, T2, and control (C) groups, as detailed below in our primary reports to SSA (see Chapter X). Broadly, we will estimate impacts of the benefit offset by examining T1 compared to the control group and T2 compared to the control group. We will estimate differences in impacts for the two treatment approaches by comparing the T1 and T2 groups. The estimation approaches for each of these comparisons are identical, differing only in which pair of groups is included. We also present some alternative approaches to estimating impacts that we could explore as a special topics for SSA. Below, we highlight some special topics to potentially complement our base analysis for these special topic briefs.

A. Research questions and data sources

We differentiate between the primary and secondary outcomes to distinguish the measures that should receive the most policy focus in the ultimate evaluation of the benefit offset's efficacy (Exhibit VIII.1). This is a transparent way to avoid concerns about data mining when assessing impacts on the broad range of outcomes (which we discuss further below). This approach also aligns with the presentation of primary and secondary research questions for the POD impact analysis.

Exhibit VIII.1. Impact analysis research questions

Domain	Question	Data source			
		Program data	Management information system (MIS) data	Survey data	Qualitative data
Primary					
Employment	Does the benefit offset increase the number of beneficiaries employed at substantive levels, defined as earnings greater than the Substantial Gainful Activity level?	X			
Earnings	Does the benefit offset increase earnings?	X			
Income	Does the benefit offset increase mean beneficiary income (earnings plus benefits)?	X			
Benefit receipt	Does the benefit offset reduce average benefits?	X			
Secondary					
Employment	Does POD affect the employment rate?	X		X	
	Does POD affect how many subjects have jobs offering fringe benefits (like paid vacation or health insurance)?			X	
	Was there a change in job search activities?			X	
Earnings	Did monthly earnings change?			X	
	Did earnings exceed Trial Work Period amounts or multiples of substantial gainful activity (2 x SGA and 3 x SGA)	X			
Income	Do benefits replace or supplement earnings?	X			
	Does the benefit offset change total income, including the use of other income supports?	X		X	
Benefit receipt	Does the benefit offset change the frequency of benefit suspensions and/or terminations?	X	X		
	Does the benefit offset change monthly variations in benefit amounts?	X			
	Does the duration of SSDI payments change?	X			
	Is the benefit offset related to a change in overpayments?	X			
	Do SSI payments change?	X			
Other	Does use of other programs, such as Vocational Rehabilitation, change?	X		X	
	Did the offset affect other health related outcomes, including physical and mental health status, hospitalization, or current health insurance?			X	
	Was there a change in job training experience and education in the last 12 months?			X	
	Does health status or quality of life change?			X	

Notes: See Chapter IV for the specific data sources included in the program, MIS, and survey data. See Chapter V for the specific data sources included in the qualitative data.

1. Primary research outcomes

We will construct the four primary outcomes by calendar year using SSA program data. We use an annual definition of each measure because our substantive employment and earnings measures, both of which are based on data from the IRS, are only available on an annual basis (see Chapter IV). The primary outcomes are all observed in program data, which means that none will suffer from potential bias arising from survey nonresponse (or misreporting by the subject). This may be especially important for earnings; Wittenburg et al. (2016) indicate administrative measures of earnings for SSDI survey respondents are, on average, higher than those respondents report to the interviewer. Additionally, we can compare these outcomes for POD subjects to comparable outcomes for nonsubjects, regardless of whether the nonsubjects were in contact by the demonstration. Finally, SSA can continue to track these outcomes after the demonstration ends.

2. Secondary research outcomes

We will also construct several secondary outcomes to address the research questions, across five domains (employment, earnings, income, benefit receipt, and other). The first four domains correspond to the substantive areas addressed by the four primary outcomes above. The fifth domain includes other outcomes that are especially important to the beneficiary, such as health, that the offset could potentially influence if it has substantive impacts on the primary outcomes. The outcomes in the five domains include administrative outcomes related to all of the secondary research questions about employment, benefit receipt, and income displayed above, as well as alternative measures of substantive employment (for example, earnings above the POD threshold amount). These secondary outcomes also include survey-based measures (see Chapter IV) of the benefit offset's potential effects in several domains, including employment (such as receipt of job training); earnings (such as changes in annual earnings in administrative data); income (such as total family income, including all other earnings and programs); and other outcomes (such as health and use of health insurance).

B. Analytic approach for overall impacts

In this section we detail our analytical approach for estimating impacts of the benefit offset. We start by describing diagnostic analyses we will conduct for the analysis sample followed by the empirical model we will estimate for the study sample.

1. Diagnostic analyses

Baseline balance testing. As a starting point for producing impact estimates, we will assess the balance of baseline characteristics of all POD participant groups (Exhibit VIII.2) It is important to note that we can also supplement the data in the exhibit with characteristics from the baseline survey. We did not do so here because there will be a lag in obtaining the baseline survey data, whereas the administrative data are regularly updated. The exhibit below shows an example of a table that we will present to SSA to provide regular updates on equivalence testing throughout recruitment. In our formal reports to SSA, we will include both administrative and baseline survey variables for the testing, and we note that we will use both administrative and survey variables later in the analytic processes as control variables and in constructing balance weights.

Our stratified, individual-level random assignment process should produce treatment and control groups that are essentially similar on baseline characteristics—that is, statistically balanced at baseline. To assess whether overall baseline covariate differences between the treatment and control groups exceed what we would expect by chance, we will conduct an omnibus test. The baseline balance assessments will be conducted for all three comparisons: T1 to C, T2 to C, and T1 to T2. To assess the size of variable-by-variable imbalance, we will calculate standardized differences and conduct t-tests. For the standardized differences analysis, we will calculate differences in variable means between experimental groups, scale each difference by the variable's standard deviation (creating a standardized difference), and then note differences that exceed 0.25 of the standard deviation. Differences of 0.25 standard deviation or greater are of greater risk of not being adequately adjusted in the regression framework (Rubin 1973).

Exhibit VIII.2. Baseline covariate balance by experimental group

Variable	Sample mean			Standardized differences		
	T1	T2	C	T1 vs. C	T2 vs. C	T1 vs. T2
Number of beneficiaries						
Gender (percent)						
Male	nn.n%			nn.n (nn.n)		
Female						
Age (percent)						
20–29 years						
30–39 years						
40–44 years						
45–49 years						
50–54 years						
55–59 years						
Mean age (years)						
Other background characteristics (percent)						
Completed high school						
Non-white						
Number of dependents and auxiliaries in SSDI						
Living in poverty						
Health status is poor						
Primary impairment (percent)						
Neoplasms						
Mental disorders						
Back or other musculoskeletal						
Nervous system disorders						
Circulatory system disorders						
Genitourinary system disorders						
Injuries						
Respiratory						
Severe visual impairments						
Digestive system						
Other impairments						
Unknown impairments						

EXHIBIT VIII.2 (*continued*)

Variable	Sample mean			Standardized differences		
	T1	T2	C	T1 vs. C	T2 vs. C	T1 vs. T2
Years since onset of disability (percent)						
Fewer than 2 years						
2 to fewer than 4 years						
4 to fewer than 6 years						
6 to fewer than 8 years						
8 to fewer than 10 years						
10 to fewer than 12 years						
12 years or greater						
Recent work and SSDI history						
Recent annual earnings (2016)						
Earned $1,000 or more in at least one of past 12 months						
Completion of TWP						
Completion of Grace Period						
Monthly SSDI Benefits ($)						
Number of months received SSDI						
SSDI only (percent)						
Concurrent beneficiary (percent)						
F Statistic						

Note: Variables in ***bold italics*** will be used to stratify random assignment, so we expect close balance on those variables. See Chapter III for details on the stratification procedures. The variables here are almost all from SSA administrative data, and we also plan to add data elements from the survey (such as education and whether earned $1,000 or more in at least one month) in our reports once those data become available.

Source: SSA program data and baseline survey.

Item nonresponse. The baseline and follow-up surveys may have some item nonresponse on particular questions. Our past experience suggests that item-level nonresponse will be low for the baseline survey, although some item nonresponse is inevitable. To address this when running subsequent analyses that draw on the baseline survey, we will set missing values in the data to a special constant value to indicate that they are missing and include a missing value indicator variable in the analyses. Likewise, although we expect that rates of item nonresponse on the follow-up survey will be very low, some item nonresponse is inevitable. The follow-up survey will primarily collect data on outcome measures to be used in the impact analysis. We believe there is little risk of systematic missing data for the outcomes in the follow-up survey. Hence, once confirming that rates of missing data are low and not systematically missing, we will omit subjects who are missing data on a given outcome from the analysis of that outcome because doing so should not bias the impact estimate.

2. Empirical model

Our analysis will focus on intent-to-treat estimates, which measure how the *offer* of the benefit offset shaped subjects' behavior after they volunteered for the demonstration. The intent-to-treat framework implies that we will not omit from the impact analysis those who withdraw from POD and revert to current law.

We can use the simple difference in mean outcomes between the treatment and control groups to estimate the intent-to-treat impacts of the benefit offset. These descriptive statistics are a straightforward and intuitive way to measure the causal effects of the new offset and

suspense/termination policies on beneficiaries who had access to them. We will use weights to account for treatment-control imbalances observed in the diagnostic analyses, constructing weights as discussed in the next section. However, weights might not eliminate all imbalances, particularly for samples with rare values of the covariates. Adjusting for baseline characteristics in a regression model after weighting is a second defense against potential bias that might arise (Robins et al. 2007). In addition, adjusting for covariates using regression models can increase the precision of impact estimates.

Our main impact model will be a weighted linear regression model that pools data from all states (unless we find that some states need to be excluded because of differences in implementation or confounding contextual factors) and includes interactions to allow for state-level heterogeneity in beneficiary characteristics, program implementation, and contextual factors.[18] We will use the following interacted regression model to estimate the pooled impact of each POD treatment arm:

$$[1] \qquad y_{is} = c_s + d_1 T1_i + d_2 T2_i + \boldsymbol{b}_s' \boldsymbol{X}_i + u_{is}$$

where y_{is} is the outcome of interest for individual i in state s, c_s is a state fixed effect, $T1_i$ and $T2_i$ denote assignment to the two POD treatment groups, $\boldsymbol{X}_i$ denotes the set of K covariates, and u_{is} is an error term.

We plan to include all covariates indicated in Exhibit VIII.2. The variables are in the list because of their policy relevance and expected predictive power. We will also present results estimated as simple differences to verify that the weights and regression adjustment themselves are improving precision and balance, as intended, but not substantively changing the findings.

The results from the linear regression model analysis will have an immediate and straightforward interpretation for understanding the effects of the benefit offset. Estimates of d_1 and d_2 represent the intent-to-treat impacts of being assigned to each treatment arm relative to the control group; we will mean-center the covariates so that a weighted average of the estimated state fixed effects, *c,* is readily interpretable as the mean level of the outcome of the control group. We can also use these estimates transparently to produce adjusted means for the two treatment groups (based on $c + d_1$ and $c + d_2$, respectively) to facilitate a graphical representation suitable for a wide audience. In contrast, nonlinear models such as logit produce estimates that (1) are less immediately interpretable and (2) tend to closely align with results from a linear model when converted into more meaningful impact estimates (Wooldridge 2010). However, for each binary outcome that addresses a primary research question, we will estimate a logistic regression model to verify that the logistic and linear regression model estimates have the same direction and significance level as well as to measure the difference between the two estimates.

[18] In addition to the main analysis, we will analyze outcomes for any states that need to be excluded from the main analysis because of differences in implementation or confounding contextual factors. For those states, we would estimate a separate set of regressions using the same model [1]; dropping state specific terms from [1] if the separate set of regressions included just one state.

Because SSA and the implementation contractor chose the POD states selectively and volunteers who ultimately enroll in the demonstration will not be a representative random sample of the population of beneficiaries from whom they were recruited, our estimates of precision will focus on inference about the baseline sample of subjects who were randomly assigned. The impact estimates $\hat{d}_1$ and $\hat{d}_2$ obtained from Equation 1 correspond to what would be obtained if we were to instead (1) estimate separate, state-specific impacts and (2) construct a weighed mean, with weights proportional to the number of study subjects in each state, to produce a pooled impact estimate (Raudenbush and Bloom 2015).

Balance weights. We will use the results of the baseline equivalence tests to develop a set of balance weights to use for the impact analysis. Balance weights can be an important complement to regression adjustment when seeking unbiased impact estimates that are robust to imbalances between groups, particularly when considering low-prevalence characteristics that can only be imprecisely controlled for in a regression (Robins et al. 2007). Additionally, covariate imbalance for analysis of survey-based outcomes can necessitate the use of weights to address issues with survey nonresponse. Consequently, we will develop two sets of balance weights: one for analysis of outcomes captured in program data that only accounts for covariate imbalance due to random assignment and a second for analysis of outcomes captured in survey data that also accounts for survey nonresponse. We will estimate balance weights using propensity scores developed from logistic regression models that we will fit using a rich set of covariates available from the administrative and survey data. Propensity scores provide a concise way to summarize and correct for initial imbalances (Imbens and Rubin 2015; Särndal et al. 1992). Given these rich data, we will use a mixture of substantive knowledge and automated machine learning methods to identify covariates to include in the final weights. In particular, we will use prespecified decision rules, such as those described by Biggs et al. (1991) and Imbens and Rubin (2015), to identify balance models that best make use of the potentially quite large number of covariates from which to choose.

C. Subgroup analyses

Heterogeneity in POD's effects across subjects could have policy implications that are of interest to SSA and other stakeholders, so it is important to measure variation in outcomes across subgroups. Focusing on a select set of subgroups will facilitate clear, interpretable, and reliable findings about how POD's impacts vary across different types of beneficiaries.

We will measure subgroup impacts for primary and secondary outcomes and report findings for subgroups in the same way we report findings for secondary research questions. Specifically, we will report findings for all subgroups using tabular structures similar to other outcomes (see below). However, we will place more emphasis on the interpretation of primary outcomes for the entire sample than on findings for any particular subgroup, even for subgroup impacts of primary outcomes.

We plan to use base subgroups on SSI status and duration of SSDI benefits from the BOND evaluation. The use of the same subgroups in POD as we used in BOND will facilitate comparisons between the BOND and POD evaluations, comparisons that are likely to be of value to SSA, given both test an offset (see Bell et al. 2011 for more details about the justification for SSI status and SSDI duration).

In a special topics reports, we could utilize methods to examine correlations between subgroups or outcomes for smaller subgroups using alternative estimation approaches. For example, one alternative is to employ a Bayesian approach to supplement the analysis above, which would address potential limitations in sample sizes (see text box and Appendix E for more details).

Special topic: Alternative options to estimate subgroups

The Bayesian analysis framework is well suited for addressing subgroups with a relatively small number of subjects (such as beneficiaries with a specific impairment type). For example, this might be useful in addressing whether impacts for mutually exclusive subgroups (such as older versus younger beneficiaries) are correlated. Another potentially useful feature of the Bayesian subgroup analysis is that we can use it to make more nuanced statements about the benefit offset's likely effectiveness for specific types of beneficiaries. For example, we could calculate the probability that the benefit offset had large impacts for younger beneficiaries, say, based on the assumption that there is some correlation across age-specific impacts.

Results from our process and participation analyses might also allow us to identify additional subgroups or refine our subgroup definitions. One particularly promising approach is to define subgroups based on whether they have begun the Grace Period at baseline. As noted in Chapters II and III, the incentives for treatment and control subjects fundamentally change based on their Grace Period status. We will use our participation analysis to assess the potential size of this subgroup if sufficient sample sizes exist (see Exhibit VIII.4). Similarly, if our process findings show substantial state differences, we might want to review state subgroups. Based on these factors, we will develop, with SSA's agreement, a final list of subgroups defined by location, beneficiary characteristics, or both to analyze in the impact evaluation. In Exhibit VIII.3, we highlight other possible subgroups that could emerge from the process and participation analyses (including the two aforementioned groups).

Exhibit VIII.3. Baseline characteristics defining possible POD subgroups

Subgroup
Beneficiaries with short (fewer than three years) versus long tenures
Beneficiaries younger than age 45 versus those age 45 and older
SSI status
Beneficiaries with different impairment types
Beneficiaries who have already assigned tickets versus those who did not
Employed versus not employed at baseline
Beneficiaries with earnings greater than the annualized SGA amount versus those who earned less than that amount during the calendar year before volunteering
Beneficiaries who have begun the Grace Period at baseline
Beneficiaries by state of residence

D. Accounting for treatment group withdrawals

As noted in Chapter VII, subjects withdrawing from the POD study could be a concern that would influence the impact estimates. If a consequential proportion of the T1 or T2 groups withdraws from POD early in the evaluation period, there will still likely be interest in understanding the impacts on those who did not withdraw.

We will use our participation analysis to guide our analysis of withdrawals as the problem of withdrawals is a combination of both the number of withdrawals and the timing of withdrawals. For example, if a large number of people withdraw at the beginning of the demonstration, the "dosage" of the intervention (that is, the length of the exposure to the offset) is limited.

Any analysis of the effects of withdrawal rates will be exploratory because subjects self-select into withdrawals and we expect that treatment group subjects will withdraw at higher rates than control subjects. This self-selection indicates that dropping withdrawals would compromise the validity of the estimates. For this reason, we would need to rely on non-experimental methods to make adjustments for withdrawal rates. One straightforward option is to use a propensity score matching method to estimate impacts for the subset of treatment subjects who remain in the study sufficiently long to be potentially affected by the new POD rules. To implement this approach, we would identify a subset of treatment subjects (for T1 and T2 separately or combined) and a set of control subjects with similar baseline characteristics, and then estimate impacts for those subsamples. We would match subjects using the rich set of baseline covariates from SSA administrative data using methods that have been previously tested with matching in the literature (see, for example, Cook, Shadish, and Wong 2008). Following this literature, we would demonstrate that the matches from the propensity score methods was successful in balancing the observable characteristics of treatment and control group subjects before presenting any comparisons.

This analysis complements our planned approach in the participation analysis, where we seek to understand why subjects were withdrawing from the POD rules (see Chapter VII, Section B.2). For example, we would examine whether subjects who withdraw are those who have used the benefit offset, and for subjects in the T2 group, whether they face benefit termination for having reached full offset. The combined analysis would provide insights into whether those who withdraw are actually subjects most affected by the offset.

E. Statistical power and precision

We expect that the sample size of eligible study subjects will be sufficient to support an analysis that can reliably detect meaningful overall impacts of the benefit offset with high probability. Calculating minimum detectable impacts (MDIs) is a standard way to characterize the expected precision of the evaluation's results, given the sample sizes and research design.

MDIs quantify the smallest true impact that is likely to be significantly different from zero, based on a two-sided statistical test of differences. We have calculated MDIs separately for outcomes measured in the SSA program data for all study subjects and outcomes measured for the subsets of subjects who complete the two follow-up surveys. We have also produced MDIs for

subgroups of varying sizes as well as for single states assumed to have differing numbers of study subjects.[19]

Exhibit VIII.4. Minimum detectable impacts for POD evaluation

Group/subgroup	Annual earnings > 12 × SGA (SSA program data)	Searching for work at time of first survey (interim survey, sample)	Searching for work at time of second survey (final survey, 100% sample)
MDIs for pairwise comparison of two study groups (percentage points)			
All beneficiaries	2.0	2.6	1.8
66 percent subgroup	2.5	3.2	2.2
50 percent subgroup	2.8	3.7	2.6
33 percent subgroup	3.5	4.5	3.2
Beneficiaries in a large state	4.5	5.8	4.1
Beneficiaries in a medium state	6.3	8.2	5.8
Beneficiaries in a small state	11.0	14.2	10.0
MDIs for comparison of both treatment groups combined vs. control group (percentage points)			
All beneficiaries	1.7	2.7	1.9
66 percent subgroup	2.1	3.4	2.4
50 percent subgroup	2.5	3.9	2.7
33 percent subgroup	3.0	4.8	3.4
Beneficiaries in a large state	3.9	6.1	4.3
Beneficiaries in a medium state	5.5	8.7	6.1
Beneficiaries in a small state	9.5	15.0	10.6
Key assumptions			
Assumed outcome prevalence in the control group (percent)	15	15	15
Total sample size	**9,000**	**3,600**	**7,200**

Note: We discuss the rationale for the assumption about control-group outcomes in the main text. Additional assumptions for POD MDIs: (1) To illustrate the MDIs, we use a basis of 9,000 enrolled beneficiaries; (2) large state = 1,800 study subjects, medium state = 900 study subjects, and small state = 300 study subjects; (3) we require at least an 80 percent chance of correctly identifying true impacts as statistically significant using two-tailed statistical tests with a 5 percent significance level; (4) we will estimate impacts using regression models that include baseline covariates explaining 40 percent of the variation in employment outcomes, that is, has an R-square of 0.40; and (5) analysis weights or adjustments for heteroscedasticity will not substantially alter variance estimates. Further assumptions for the surveys are that (1) we will field the first survey to half of the study subjects, (2) we will field the second survey to all subjects, and (3) approximately 80 percent of potential respondents will complete surveys fielded to them.

MDIs for outcomes measured in SSA program data for all subjects. We expect that impacts based on SSA program data are likely to have sufficient precision to reliably detect impacts shown in the table for each of the two POD treatment arms. The first two columns of Exhibit VIII.4 include illustrative outcomes for one of the primary outcomes, substantive earnings, defined as annual earnings greater than 12 times SGA level. This outcome is available in the SSA program data and therefore available for all study subjects. Calculating these MDIs required us to make an assumption about the prevalence of each outcome in the control group. In the BOND Stage 2 control group, approximately 9 percent had substantive earnings (Gubits et al.

[19] Based on the information presented in Chapter II, Section B, the sample size assumed for a large state in Exhibit VIII.4 is roughly the number expected from Alabama, and the sample size assumed for a small state in the exhibit is between the numbers expected from Nebraska and Vermont.

2014). Because we expect POD to be more attractive to work-interested beneficiaries than BOND, we assumed a higher prevalence of substantial earnings in the POD control group, 15 percent.

We expect the overall MDI for estimated impacts on SGA-level earnings for each treatment arm to be 1.5 percentage points. This MDI suggests we will be able to reliably detect relatively small impacts for each treatment arm, compared to the control group under current law. This MDI is 10 percent the size of the assumed prevalence under current law (1.5/15 = 0.10). The MDIs for comparing the two treatment arms (combined) against the control group are slightly smaller—for example, 1.3 percentage points for the annual employment rate versus 1.5 percent when considering for each treatment group separately.

For subgroups based on beneficiary characteristics or for particular states, impacts of POD would need to be somewhat larger to be reliably detected, but we expect these subgroup analyses would still have enough precision to be informative. For example, all impacts—for both pairwise comparisons and comparisons between the two treatment groups (combined) and the control group—would have to be almost 25 percent higher to be reliably detected for a subgroup made up of 66 percent of study subjects. Based on the BOND Stage 2 sample (Gubits et al. 2014), potential examples of such large subgroups include subjects with short SSDI durations or those without a college degree. Impacts would need to be correspondingly larger for more-focused subgroups containing smaller percentages of the study subjects.

MDIs for outcomes measured using survey data for subsets of subjects. We expect to have less precision for impacts on survey-based outcome measures, given the smaller sample sizes, but are likely to be able to detect modest-sized impacts for all but the smallest subgroups. The final two columns of Exhibit VIII.4 report MDIs for the share of subjects actively searching for work at the time of each of the two follow-up surveys. Based on anticipated response rates of 80 percent, we expect a respondent sample size of 6,000 for the Year 1 survey (which we will field to half of the subjects) and 12,000 for the Year 2 survey. For comparative purposes, the prevalence of work search in the control group at both points is assumed to be 15 percent—the same as the prevalence of annual earnings greater than 12 times the SGA amount.[20] MDIs for pairwise comparisons of work search are 2.5 percentage points based on the Year 1 survey and 1.7 percentage points based on the Year 2 survey.

F. Reporting and interpreting impacts

Effectively using research evidence to inform policy requires that the evidence be clearly and accurately communicated. Our approach to reporting and interpreting impact estimates is

[20] We based this assumption about work-search rates in the POD control group on information about respondents to the National Beneficiary Survey who expressed an interest in work. Among SSDI-only beneficiaries who responded to the National Beneficiary Survey in 2010, 4 percent had searched for work over the past four weeks, and a total of 26 percent had work goals that included finding a job, improving skills, or career advancement (Wright et al. 2012). This suggests a work-search rate of approximately 15 percent (= 0.04/0.26) among work-interested respondents. Livermore (2009) found a similar rate of work search (13 percent) among beneficiaries with a work orientation who responded to the 2004 National Beneficiary Survey.

intended to facilitate an accurate understanding of the likelihood that the POD design truly has effects and the substantive importance of those effects.

1. Reporting impacts

We plan to report regression-adjusted mean values for each outcome and to calculate impact estimates, standard errors, and associated *p*-values. Impacts will be estimated separately between each combination of experimental groups. As an example, Exhibit VIII.5 shows how impact estimates for the primary outcomes will be displayed numerically in a more detailed format for appendix tables. For the main text of the reports, we plan to describe the primary outcome estimates graphically using simple bar charts that will be accessible for a broad audience.

Exhibit VIII.5. Impacts for primary outcome measures

	Sample mean			Impact estimate		
Outcome	T1	T2	C	T1 vs. C	T2 vs. C	T1 vs. T2
Substantive employment (percent with annual earnings about annualized SGA)	nn.n%			nn.n (nn.n)		
Earnings ($)						
Annual SSDI benefit amount ($)						
Total annual income: SSDI benefits plus earnings ($)						

Source: SSA program data.

Notes: Standard errors are reported in parentheses. */**/*** indicates statistical significance at the .10/.05/.01 level.

Exhibit VIII.6 lists other administrative outcomes for the secondary research questions as well as outcomes derived from survey data and state program data sources from the following five domains (employment, earnings, benefit receipt, income, and other). These outcomes span the same four domains as the primary outcomes above plus a fifth (other) domain. Our analysis of these secondary outcomes will provide additional exploration of the factors that might be driving the effects. For example, if we find impacts on earnings, our subgroup impacts on different levels of earnings (such as earnings below SGA and earnings two times SGA) will provide context for the factors driving the earnings impacts. Our analyses of the other domains provide insight into the offset's impacts on other factors that might be related to our primary outcomes, such as job search in the case of earnings and health in the case of all outcomes. As they will in reporting primary outcomes, our reports will present the secondary outcome findings graphically, with detailed estimates in appendices.

2. Multiple comparisons and sensitivity analyses

Assessing whether a statistically significant impact estimate of the benefit offset is due to a true program effect rather than random chance requires more information than our estimated impact and *p*-value. A common mistake is to interpret the *p*-value as the probability that the true impact is zero, given what we observe in our data (or, equivalently, that the estimated impact is due to randomness alone). In 2016, the American Statistical Association issued a statement explaining the consequences of this misinterpretation of *p*-values. This misinterpretation has contributed to the "replication crisis" in many fields because the probability that the impact is close to zero is often much higher than the *p*-value (Wasserstein and Lazar 2016; Greenland et

al. 2016). By replication crisis, we mean that many statistically significant findings cannot be replicated in subsequent studies, suggesting that the original findings might have actually been due to random chance rather than true effects.

The misinterpretation of *p*-values, and the associated replication crisis, can be thought of as a problem of multiple hypothesis testing. When multiple hypotheses are tested within a study, the false discovery rate—that is, the proportion of statistically significant impacts that are due to random chance, not a true program effect—can be much greater than the level of significance (typically 5 percent) used in testing.[21] The replication crisis exists because this problem also exists *across* studies. To distinguish true impacts, our interpretation of impacts will focus on assessing (1) whether the estimated impacts are likely to represent true program effects rather than chance differences between the treatment and control groups and (2) whether the program impacts are of a substantively important magnitude.

Our main approach for mitigating multiple comparison issues is to specify a small number (four) of primary outcomes, which we will use to determine if the benefit offset has important impacts as discussed in the previous section. Specifically, in interpreting findings for our reports, we will place more emphasis on the interpretation of primary outcomes than of secondary outcomes. For example, in our executive summaries, we will always present findings for our four primary outcomes in all of our planned reports that include impact estimates (see Chapter X for more details).

Another way of guarding against misinterpreting findings from the study is to look for patterns of results that are consistent with the theory underlying the intervention. From a policy perspective, the most important impacts will be those on substantial employment, earnings, benefits, and income. Our assessment of whether the magnitude of program effects is substantively important will be based on its magnitude relative to other key benefits and costs in the benefit-cost analysis, as described in Chapter IX.

An additional sensitivity test is to draw on findings from other studies to formally assess the probability that a statistically significant impact estimate is due to a true program effect rather than random chance. This approach provides an alternative solution for the issue for which *p*-value adjustments for multiple comparisons have been used in the past. Our plan would be to show these estimates and charts formally in the appendix and only reference them in the reports' main text if they substantively change our interpretation. In Appendix C, we show how we would show these sensitivity tests using a set of charts. Operationally, we would only cite findings from these charts in reporting impacts if there is a substantive multiple comparisons problem. In the appendix, we also note how we would bring this information into the presentation of findings as a sensitivity test to make adjustments for the final estimates.

[21] The false discovery rate could also be *less than* the level of statistical significance used in testing. This would be the case if more than half of interventions were truly effective. More details on calculating the false discovery rate are provided in Appendix E.

Exhibit VIII.6. Impacts for secondary outcome measures

Outcome	Sample mean			Impact estimate		
	T1	T2	C	T1 vs. C	T2 vs. C	T1 vs. T2
Employment						
Actively searching for job (percent)						
Any employment (percent)						
Hours worked per week at current or most recent job						
Current or most recent job offered [paid time off/health insurance/other specific fringe benefits] (percent)						
Current or most recent employer has made accommodations for physical or mental conditions (percent)						
Earnings						
Earning more than two times the annualized SGA amount (percent)						
Earning more than three times the annualized SGA amount (percent)						
Earnings at current or most recent job ($)						
Benefit receipt						
SSDI benefit months (months)						
SSDI suspension or termination months (months)						
Annual SSI receipt amount ($)						
SSI benefit months (months)						
SSI suspension or termination months (months)						
Income						
Income from [veteran's benefits/public assistance/worker's compensation/unemployment/food assistance/other specific social programs]						
Total family income ($)						
Other						
Health condition is [poor/fair/good/very good/excellent]						
Regular daily activities limited by [physical/emotional problems/pain] (percent)						
Other physical and mental well-being indicators						
Has health insurance (by specific type)						
Personal goals include getting job or improving job skills						
Personal goals include earning enough to not receive SSDI						

Source: SSA program and earnings administrative data and survey data from the Year 1 and 2 follow-up surveys.

Notes: Standard errors are reported in parentheses.

G. Extrapolating results for a national program

We plan to use information from the impact analysis in combination with the process analysis to make inferences about national policies using simple simulations. This information will be useful in understanding the national implications from the evaluation.

There are two fundamental challenges to extrapolating the results. First, as discussed in Chapter II, a fundamental challenge to extrapolating results is the specification that SSA operate demonstrations that include volunteers who provide written consent and can withdraw at any point. Second, the eight states included in the demonstration, while regionally diverse, were not sampled in a manner to produce national estimates (see Chapter III). Another consideration is whether the extrapolation should be based on a voluntary or mandatory version of POD as a national policy. If voluntary, similar to the rules of POD, beneficiaries could opt into one of the two POD offset designs or stay with the current rules. If mandatory, the POD offset rules would replace the current rules for all beneficiaries.

For voluntary versions of the policy, we would assume a national policy would replicate the patterns observed from the participation and impact analysis. For example, we would link the characteristics of volunteers in our demonstration sample to a national sample to project a volunteer rate. Specifically, we would use realignment weights that adjust the balance weights (described above in VIII.B.2) so that the weighted distribution of characteristics in the study sample matches the larger population of beneficiaries that would be expected to use the offset. For that group of volunteers, we would then apply the impact estimates we observe for volunteers to the full national sample to extrapolate the national findings. We would develop these weights using population characteristics from program data and rates of volunteering for POD, focusing on those recruited from the samples randomly selected for outreach and recruitment efforts. These simple simulations would provide a general estimate for a national policy, though we could test further assumptions to illustrate the sensitivity of these estimates.

Special topic options for testing generalizing assumptions

- **Volunteer rates:** We could use differential rates of volunteering for POD to simulate national differences in participation.
- **Beneficiary characteristics:** We could test variations in realignment weights that assign less weight to individuals in the group that is overrepresented and more weight to individuals in the group that is underrepresented.
- **Distribution of impacts:** We could test variations in outcomes based on the distribution of impacts by subgroups and reassign based on beneficiary characteristics.

For the mandatory version, we need to apply estimates of POD use to volunteers and make strong assumptions for non-volunteers who would eventually use POD. The starting point for these estimates would include the volunteer methodology noted above to simulate POD's impact for that group of beneficiaries. We would then need to make assumptions about behavior for non-volunteers who might use POD under a national program, which we do not observe in the demonstration. Nonetheless, we could make some reasonable assumptions as a starting point. For example, a simple assumption is to include beneficiaries who have earnings above the TWP who were not present in the volunteer sample above. This assumption would represent a lower bound of estimates, given that some beneficiaries who are now working at levels above TWP would presumably also eventually use POD as they learned about the new rules.

In presenting the results for both the voluntary and mandatory options, we would caveat our findings about other factors that could heavily influence the sensitivity of the extrapolation. For example, in both the voluntary and mandatory versions of the policy, we would have to consider whether certain parameters of the demonstration authority would still apply, such as the need for written consent. In reporting our findings, we will clearly list these assumptions to provide the readers the important caveats necessary to extrapolate the results to national policy given the two demonstration challenges noted above. We will work with SSA to identify the appropriate assumptions for extrapolation, which we anticipate will create ranges for the projections to national policy. We will report these findings in special topic reports.

IX. BENEFIT-COST ANALYSIS

The benefit-cost analysis will assess:

- ***The benefits of the POD benefit offset and associated services net of costs.*** Was the benefit offset and its associated components cost beneficial when considering the combined benefits and costs to all beneficiaries involved in the demonstration and to government entities? Was the benefit offset cost beneficial to particular groups—particularly beneficiaries and SSA?

We will summarize our approach using an accounting framework that draws on internal and external data sources. We will use these sources to place a dollar value on each benefit and cost of the intervention. We will then combine all of the monetized costs and benefits into one statistic, the net benefit, which summarizes the benefit offset's potentially diverse benefits and costs. For example, we can use estimates from the impact analysis to quantify many of the benefits (such as employment and earnings) and costs (such as changes in benefits). However, to assign a full monetary value of benefits and costs, such as implications for tax revenues, we must also obtain information from outside resources. Our approach builds off of the benefit-cost analysis conducted in other evaluations, including BOND (Bell et al. 2011), YTD (Fraker et al. 2014), the Workforce Investment Act evaluation (Fortson et al. 2017), and the Individual Training Account Experiment (McConnell et al. 2006).

A. Accounting framework

We will develop a comprehensive accounting framework that incorporates a range of perspectives to guide benefit-cost data collection, analysis, and reporting. In Exhibit IX.1, we show our planned framework for the analysis, reflecting the stakeholder perspectives that we propose to include, along with the potential financial and nonfinancial impacts of the treatments. Line items in the exhibit indicate various components of benefits and costs, and we note the expected monetized value of the change for each entry in the framework. The Benefits noted in a stakeholder column indicate projected sources of benefits for the stakeholder, the Costs indicate projected sources of costs for the stakeholder, zeros (0) imply no projected effect, and question marks (?) indicate no predicted sign for the impact. For example, we expect that SSI benefits will be lower for treatment group subjects, which implies a monetary cost for treatment group subjects but a monetary benefit for the other federal agencies, and no change for the other stakeholders. We will present the benefit-cost analysis in exhibits similar to Exhibit IX.1, with dollar values in each cell. The bottom row for each column sums the positive benefits minus negative costs and indicates total net benefits (or total net costs) to the stakeholder group listed in the column, in dollars.

Because some benefits and costs occur at different times, we must make two adjustments when we aggregate them. First, with inflation, a dollar in the baseline year can purchase more goods and services than a dollar in a later year. We will use the implicit gross domestic product price deflator to convert all benefits and costs occurring in later years into constant dollars. Second, even in the absence of inflation, a dollar in the baseline year is worth more than a dollar in a later year because it can be invested and earn interest. To account for this effect, we will use a discount rate to convert all future benefits and costs to their present value. Previous evaluations of social programs have used real discount rates ranging from 3 to 10 percent. As with the BOND evaluation, we propose to set the discount rate equal to the rate SSA used in its actuarial

projections of the SSDI Trust Fund balance, but we will conduct sensitivity analyses to determine how much this assumption affects the findings of benefit-cost analysis.

The analysis will assess the net benefits of the treatments on treatment group subjects, the government, and society as a whole. We discuss each perspective in turn:

- ***Treatment group subjects.*** The two treatment approaches are intended to increase earnings and substantive employment for treatment group subjects relative to the control group and, correspondingly, reduce SSDI benefits. Because beneficiaries are not likely to volunteer for POD if the offset is not financially attractive to them, the combined effect of the increased earnings and lower SSDI benefits will likely increase total income and improve other aspects of the lives of beneficiaries and their families. Subjects will generally not pay for the treatment services such as counseling. On average, however, they are likely to incur substantial costs in terms of forgone SSDI, SSI, Medicare, Medicaid, and other benefits. In addition, higher earnings will mean increased receipt of fringe benefits and higher taxes. We will use updated estimates of fringe benefits as a percentage of earnings from the Department of Labor's National Compensation Survey. We plan to use simple imputations of taxes based on average tax rates for individuals earning at levels consistent with those of the study subjects.[22]
- ***SSDI Trust Fund.*** We will conduct separate analyses for benefits and costs accruing to government entities, and can produce estimates for the SSDI Trust Fund separately from all other net benefits to the government (as shown in Exhibit IX.1).
- ***Other federal, state, and local government.*** Potential benefits to the other federal government funding include possibly significant reductions in SSI, and Medicare/Medicaid program and administrative expenditures; smaller reductions in benefits for other programs (for example, under the Supplemental Nutrition Assistance Program [SNAP]); and increased payroll tax and income tax revenues. Conversely, the benefit offset could lead to increased use of Ticket to Work services and associated costs. Besides the direct costs associated with a beneficiary's receipt of these services, we will also use the framework described in Isaacs (2008) to project the additional administrative costs associated with each of these programs, updating information on specific administrative costs from the U.S. House of Representatives' Committee on Ways and Means. We will collect data on subjects' time using POD-related services through the study survey. We will also collect cost data from states on the average hourly compensation rates of program staff to help estimate the value of the time subjects spend receiving services, as discussed in Chapter V.
- ***Society.*** Aggregating data across all perspectives will yield an estimate of net benefits to society. Benefits for stakeholders that are costs to other stakeholders will net out to zero when considering society as a whole. Positive net benefits would indicate that the program increased the value of resources available to society, regardless of which entities accrue benefits and costs; that is, the value of the increase in the productivity of treatment subjects,

[22] We considered more sophisticated approaches for imputing taxes paid, such as the National Bureau of Economic Research's TaxSim program, but given the relatively low level of earnings for the beneficiary population, we do not believe more advanced methods would be worthwhile.

represented by increases in wages and other compensation, outweighs the marginal investment in resources used to help subjects earn more.

The benefit-cost analysis will focus on impacts that may be easily monetized. It is not feasible to monetize the value of other benefits and costs, such as quality of life or social integration stemming from greater employment, except possibly in controversial ways. Therefore, we will simply present the impacts for these outcomes in tandem with the net benefit estimates and provide a qualitative assessment of their contribution to overall net benefits. The goal is to provide decision makers with a comprehensive picture of benefits and costs, even if we do not capture all elements in dollars.

Exhibit IX.1. Benefit-cost accounting framework for final report

Benefits and costs	Data source	Treatment group subjects	SSDI trust fund	Other federal	Society
SSDI benefits	SSA	?	?	0	0
SSI benefits	SSA	Cost	0	Benefit	0
SSDI administrative costs	SSA	0	Benefit	0	Benefit
SSI administrative costs	SSA	0	0	Benefit	Benefit
Ticket to Work costs	SSA	0	Cost	0	Cost
Medicare/Medicaid payments	House Committee on Ways and Means	Cost	0	Benefit	0
Medicare/Medicaid administrative costs	House Committee on Ways and Means	0	0	?	?
Earnings and fringe benefits	SSA and National Compensation Survey	?	?	?	?
Payroll taxes	Imputed based on earnings	Cost	Benefit	Benefit	0
Income and sales taxes	Imputed based on earnings	Cost	0	Benefit	0
Unemployment Insurance (UI) benefits	Survey	Benefit	0	Cost	0
Other programs (Temporary Assistance for Needy Families (TANF) payments and SNAP (food stamp benefits)	Survey	Cost	0	Benefit	0
POD administrative costs	SSA	0	Cost	0	Cost
Work-related expenses (child care, transportation, clothing)	Survey	Cost	0	0	Cost
Quality of life and well-being	Survey	Benefit	0	0	Benefit
Net benefits / costs		**?**	**?**	**?**	**?**

Note: The signs in the table indicates the monetized value of what the impact contributes to the overall benefit-cost calculation. Specifically, Benefit indicates a predicted monetary benefit for the stakeholder based on the expected change for the measure; Cost indicates a predicted monetary cost for the stakeholder based on the expected change for the measure; 0 indicates a prediction of no monetary change for the stakeholder based on the expected change for the measure; and ? indicates uncertain change the stakeholder for the measure.

B. Special options for adjusting accounting framework

Special Benefit Cost Analysis topics

- **Projecting future impacts.** The benefit cost framework above represents a standard set of assumptions that our team could test using alternative assumptions for benefit and costs outcomes (such as annual decreases and increases of 10 percent in the size of the impacts). Additionally, we could summarize the size of future effects on earnings and benefits that would be required for POD to have a positive net impact under different scenarios of a POD national policy.
- **Statistical uncertainty.** The key inputs into the benefit-cost analysis include estimated impacts with standard errors. To adjust for these issues, we could present the overall net benefits with confidence intervals using methods to adjust standard errors for key outcomes modeled as was done in a previous study (McConnell et al. 2006).

Our benefit-cost estimates in the framework above will represent estimates two to three years following random assignment, which creates an important analytic issue if we expect benefits and costs to continue to emerge beyond the evaluation period.[23] Hence, estimates of the full set of benefits and costs might depend heavily on the size of future earnings impacts, and to a lesser extent on our assumptions about the amount of future benefits received and service costs incurred. For example, Wittenburg et al. (2015) showed impacts in BOND continued to emerge through the third year of the demonstration. For this reason, SSA might be interested in special topic analyses that account for uncertainty in projected future impacts of POD. These projections could account for, say, simulations that assume longer implementation periods than allowed for the demonstration. Alternatively, SSA might be interested in options that simply project when POD might be cost neutral to the trust fund if it is not cost neutral during the period of the evaluation. Finally, SSA might be interested in testing the sensitivity of the framework to alternative statistical assumptions to test which assumptions generate cost neutrality, which might be particularly important if we have to make strong assumptions for difficult to measure outcomes noted above, such as quality of life.

[23] As noted in Chapter III, because of the rolling enrollment period, POD subjects will have different durations of participation in demonstration services. Those who enroll in the first few months of the demonstration will have approximately three years of exposure to the benefit offset and associated services, whereas those who enroll later will have approximately two years of exposure.

X. EVALUATION REPORTS AND TIMELINE

The POD evaluation reports include:

- ***Primary reports.*** We will produce four primary reports that provide updates on POD processes and outcomes.
- ***Special topic briefs.*** We will develop eight policy briefs on special topics related to the POD study and issues that emerge during the contract.

Both types of reports will provide timely and comprehensive information about the demonstration outcomes. To effectively disseminate findings, the team will write all reports in plain language intended for a general, non-technical audience.

In Exhibit X.1, we present an overview of the data content, outcome time frame, and due date for each report. We worked with SSA to identify deliverable dates that would fit within the five-year window for the demonstration, while also identifying content for each report that will provide timely findings from each of the available data sources about the progress of POD. Each report will cover a set of outcomes over the same period, which is notable given that administrative file updates on outcomes, particularly for the primary benefit and earnings outcomes, vary. As summarized in Chapter IV, SSA updates the Master Earnings File annually, whereas it updates its program data monthly. As a result, there is an approximately 10- to 11-month time lag before we can produce reports on a joint set of annual earnings and benefit outcomes for each deliverable. This time lag is consistent with the updates for evaluation reports in other SSA demonstrations.

In the remainder of this section, we provide a more in-depth summary of our plans to use the data summarized in Exhibit X.1 in each of our reports, including a tentative summary of topics. Specifically, we summarize the planned topics for each of the four primary reports and highlight possible topic areas for the eight special topic briefs.

Exhibit X.1. Deliverable schedule for evaluation reports

Deliverable	Data content	Outcome time frame	Draft due date
Draft Early Assessment Report	SSA program data RAPTER data Abt's Implementation Data System data Round 1 qualitative data	SSA program outcomes/program processes: January 2018–April 2018	10/31/2018
Draft Recruitment and Random Assignment Analysis Report	SSA program data RAPTER data Abt's Implementation Data System data Round 2 qualitative data	SSA program outcomes/program processes: January–December 2018	6/30/2019
Draft Interim Impact Evaluation Report	SSA program data RAPTER data Abt's Implementation Data System data Round 3 qualitative data VR program data IRS earnings data Baseline and available Year 1 survey	SSA program and VR outcomes/program processes: January 2018–December 2019 Earnings data: January-December 2018 Survey outcomes: through Year 1 (January 2020)	6/30/2020
Draft Final Impact Report	SSA program data RAPTER data Abt's Implementation Data System data Round 4 qualitative data VR program data IRS earnings data Year 2 survey	SSA program and VR outcomes/program processes: January 2018–December 2020 *Earnings data: January 2018-December 2019*[1] Survey outcomes: through Year 2 (January 2021)	6/30/2021
Draft Special Topics Briefs ***(8)***	TBD	TBD	TBD

Bold italics = update in contract schedule.

[1] Within one of the special topics reports we will include findings using 2020 earnings, updating the estimates in the final report.

A. Primary reports

The four primary reports will provide regular updates on the progress of POD throughout the evaluation. Our first two primary reports (an early assessment report and recruitment and random assignment analysis report) will provide an early glimpse into recruitment, enrollment, and service activities through the first year of implementation. These reports will include several of the process and participation components described in Chapters V and VI. Our second two primary reports will summarize program processes and outcomes through the end of the demonstration. Specifically, the interim impacts report will summarize findings from the planned participation, process, and impact analyses (Chapters VI–VIII). The final impacts report will update the findings from the interim report and summarize the benefit-cost analysis (Chapter IX).

Early assessment report. The early assessment report summarizes findings related to recruitment, enrollment, early service provision, and the fidelity of implementation for the first four months of the intake period. We will use a combination of program data, RAPTER data, Abt's Implementation Data System data, and Round 1 qualitative data. The early assessment report will create an opportunity to refine the enrollment processes and suggest potential areas for new or enhanced technical assistance or training for states that are not implementing services in accordance with the demonstration. We will identify any potential concerns for the internal or external validity of the evaluation, including findings from the recruitment pilot and available data comparing the characteristics of T1, T2, and C subjects to assess whether random assignment is working as envisioned. We will use the qualitative data to describe the program context, including early infrastructure development, staff involvement, and SSA operational issues (see Chapter VI). The report will identify areas for improvement in service delivery and outline challenges SSA might face when creating the operational infrastructure needed to support implementation of the offset.

Early assessment report topics

- Summary of initial recruitment and service provision
- Analysis of fidelity of implementation
- POD support infrastructure development
- Staff involvement in enrollment and implementation
- SSA operations for POD
- Recommendations for corrective action

Recruitment and random assignment analysis report. We will prepare a summary of the recruitment and random assignment report at the conclusion of the recruitment period using program data, RAPTER data, Abt's Implementation Data System data, and Round 2 qualitative data. The report will examine the overall performance of the recruitment and random assignment process, with a heavy focus on summarizing findings from the participation analysis. Because POD is using a volunteer sample, it is important to understand whether any systematic observable differences exist between POD volunteers and nonvolunteers. Differences between these two groups could have consequences for implementation, as well as the external validity of POD's findings. Therefore, the recruitment and random assignment analysis report will describe and analyze the characteristics of POD volunteers and nonvolunteers.

Recruitment and random assignment analysis report topics

- Characteristics of POD volunteers
- Characteristics of POD nonvolunteers
- Comparison of POD experimental groups
- Summary of recruitment and enrollment patterns

For those who did volunteer, we will compare and analyze differences between the treatment and control subjects. We will pay special attention to assessing baseline equivalence between the treatment and control groups and the extent to which observed imbalances between groups (if any) could bias future POD analyses. Finally, the recruitment and random assignment analysis report will draw on the findings from the second round of qualitative data collection to assess recruitment processes and to understand beneficiaries' preference for the interventions. Beneficiaries' preferences and perceptions will likely be key drivers of whether they enroll, withdraw, or use the offset. Our analysis will examine the data for patterns that will help SSA better understand what is motivating the behavior of POD subjects. Of particular interest will be understanding differences between POD subjects and the overall eligible beneficiary population in each catchment area.

Interim impacts report. The interim report will be the first report to present impacts, as well as to integrate findings from the process and participation analyses. Results from the analysis of Year 1 impacts on primary outcomes will likely draw substantial attention from policymakers interested in an initial quantitative assessment of POD. Secondary impact estimates are also likely to be of considerable interest to stakeholders.

Interim report topics

- POD states and the service environment
- Administration of the POD offset and benefits counseling services
- Beneficiaries' understanding of the offset
- Offset use in Year 1
- Primary impact estimates, Year 1
- Secondary impact estimates, Year 1

In describing the impact estimates, it will be important to put the findings into the context of how the demonstration's operations mature. This context will be based on available information from the third round of qualitative data collection and analyses of beneficiaries' participation patterns. For example, we will use information from the process analysis to identify contextual factors that might be linked to differences in impacts across states. In addition, the results of the interim impact evaluation report will provide information on whether the implementation of POD is moving toward the objectives specified in the theory of change in Chapter I, even if it has not yet achieved the long-term primary impacts. For example, following the approach we used in the evaluation of BOND, our understanding of entry into work and offset use would be informed by lessons learned from the process study about how the offset and related benefits counseling services were administered in practice. Similarly, using participation analysis results on the number of beneficiaries in the POD treatment group who understand or use the offset could help explain the size and pattern of impacts on employment and benefit outcomes.

Final impacts report. The final report will provide a summative evaluation of all demonstration activities, including whether POD met its ultimate objectives in the theory of change model described in the introduction. It will include Year 2 impacts on primary and secondary outcomes, as well as important findings in other areas. It will also synthesize contextual and administrative factors that facilitated or inhibited implementation of the offset and benefits counseling services. This information will help inform our discussion of potential system changes—for example, in benefit-processing operations, overpayment identification, and interactions with state agencies and beneficiaries—that SSA should consider if the agency were to implement the POD offset at scale. In addition, the report will include a final summary of the participation rates in all benefits counseling activities and in the offset. Furthermore, as in the interim report, we will consider how and why the demonstration produced its particular impacts based on differences identified in participation and processes across states, regions, and subgroups. The report will also supplement the main analysis with projections of impacts of a

Final report topics

- Barriers and facilitators to POD implementation
- System changes in benefit processing, program integrity, and work orientation
- Stakeholder perspectives on offset policies and benefits counseling services
- Offset use in Year 2
- Primary impact estimates, Year 2
- Secondary impact estimates, Year 2
- Benefit-cost analysis

national policy using methods described in the impact analysis. Finally, using the final round of qualitative data collection, the report will provide a systematic understanding of the perceived value of the POD offset from the perspectives of administrators, beneficiaries, and other SSA stakeholders. This summary will provide important context for the benefit-cost analysis, which will provide a summative review of POD's estimated benefits and costs to multiple stakeholders, focusing on outcomes for beneficiaries and the SSDI Trust Fund.

B. Special topic briefs

We will develop eight policy briefs on special topics related to the POD study and issues that emerge during the contract. The briefs are an opportunity to use the evidence from POD to spotlight its implications for specific SSA practices, inform future demonstration evaluations, and present results from analyses that go beyond the streamlined results presented in the impacts reports. We anticipate that topics for these briefs will crystalize as we begin to learn from the development and implementation of POD, and others will come to the forefront as the primary reports are completed. We plan to write succinct briefs that are accessible to a broad audience. This brief format naturally complements the more comprehensive format of the content included in each of the four reports described above.

Special topic briefs examples

- **Process:** Innovative program implementation practices
- **Participation:** Recruitment experiment findings
- **Impact:** Presentation of subgroup impact findings using Bayesian analyses
- **Benefit-cost:** Integrating statistical uncertainty into benefit-cost analyses
- **Cross-cutting:** Show variations of generalizations to national POD policy.

We anticipate that SSA might want some briefs that summarize the primary reports above, as well as briefs on specific topics of policy interest. In the call-out box, we list extensions of findings from the primary report from the process, participation, impact, benefit-cost and cross-cutting analyses as example topics that might be worthy of special reporting. An early special topic brief that could be of particular interest given the potential challenges in meeting recruiting targets is to summarize the findings from the recruitment experiments. This topic could be especially germane in providing an early glimpse into POD activities, as well as inform other SSA efforts to contact SSDI beneficiaries. Towards the end of the project, we anticipate a cross-cutting analysis of the demonstration implications for different versions of POD policy could be of particular interest given the challenges noted in Chapters III and VIII in developing these estimates. For example, one policy brief could examine alternative simulations of national POD policies that test different assumptions from those used in our Final Report to create bounded estimates or to test, say, the potential effects of POD in different geographic areas.

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

REFERENCES

Abt Associates Inc. "Promoting Opportunity Demonstration Implementation Design Report." Cambridge, MA: Abt Associates, 2017a.

Abt Associates Inc. "Promoting Opportunity Demonstration Site Report." Cambridge, MA: Abt Associates, 2017b.

Abt Associates Inc. "Promoting Opportunity Demonstration Technical Assistance and Site Monitoring Plan." Cambridge, MA: Abt Associates, 2017c.

Alexander, J.A., and L.R. Hearld. "Methods and Metrics Challenges of Delivery System Research." *Implementation Science,* vol. 7, no. 15, 2012. doi: 10.1186/1748-5908-7-15.

Bell, Stephen, David C. Stapleton, Daniel Gubits, David Wittenburg, Michelle Derr, David Greenberg, Arkadipta Ghosh, and Sara Ansell. "BOND Implementation and Evaluation: Evaluation Analysis Plan." Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, 2011.

Benjamini, Y. and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society Series B*, vol. 57, no. 1, 1995, pp. 289-300.

Biggs, David, Barry de Ville, and Ed Suen. "A Method of Choosing Multiway Partitions for Classification and Decision Trees." *Journal of Applied Statistics,* vol. 18, no. 1, 1991, pp. 49–62.

Borjas, George. *Labor Economics*, McGraw-Hill Education, Seventh Edition, 2016.

Colquhoun, David. "An Investigation of the False Discovery Rate and the Misinterpretation of p-Values." *Royal Society Open Science,* 2014. doi:10.1098/rsos.140216.

Cook, Thomas, William Shadish, and Vivian Wong. "Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons." *Journal of Policy Analysis and Management*, vol. 27, no. 4, 2008, pp. 724-750.

Croake, Sarah, Priyanka Anand, Christopher Jones, Katherine Morrison, Cara Orfield, David Stapleton, Denise Hoffman, David Mann, Judy Geyer, Daniel Gubits, Stephen Bell, Andrew McGuirk, Tyler Rose, and David Wittenburg. "2017 Stage 1 Interim Process, Participation, and Impact Report." Submitted to the Social Security Administration. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, 2017.

Damschroder, Laura J., David C. Aron, Rosalind E. Keith, Susan R. Kirsh, Jeffery A. Alexander, and Julie C. Lowery. "Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science." *Implementation Science,* vol. 4, 2009, pp. 50-65.

Delin, B. S., E. A. Hartman, and C. W. Sell. 2010. "Does Work Incentive Benefits Counseling Improve Employment Outcomes for Those with Serious Disabilities? Preliminary Evidence for the 'Work Oriented' from Two Demonstration Projects." *Journal of Disability Policy Studies*, vol. 26, no. 1, pp. 54-64.

Derr, Michelle, Denise Hoffman, Jillian Berk, Ann Person, David Stapleton, Sarah Croake, Christopher Jones, and Jonathan McCay. "BOND Implementation and Evaluation Process Study Report." Washington, DC: Mathematica Policy Research, February 2015.

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 Impacts Findings on the WIA Adult and Dislocated Worker Programs." Oakland, CA: Mathematica Policy Research, 2017.

Fraker, Thomas, Arif Mamun, Todd Honeycutt, Allison Thompkins, and Erin Jacobs Valentine. "Final Report on the Youth Transition Demonstration Evaluation." Washington, DC: Mathematica Policy Research, December 29, 2014.

Gelber, Alexander, Timothy J. Moore and Alexander Strand. "The Effect of Disability Insurance Payments on Beneficiaries' Earnings." *American Economic Journal: Economic Policy,* vol. 9, no. 3, 2017, pp. 229-261.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis* (Third Edition). Boca Raton, FL: Chapman & Hall/CRC Press, 2014.

Geyer, Judy, Danny Gubits, Stephen Bell, Denise Hoffman, Sarah Croake, Katie Morrison, David Judkins, and David Stapleton. "BOND Implementation and Evaluation: 2017 Stage 2 Interim Process, Participation, and Impact Report." Submitted to the Social Security Administration (contract deliverable 24c2.4 under Contract SS00-10-60011). Cambridge, MA: Abt Associates, September 2017.

Goodman, N. and Douglas G. Altman. "Statistical tests, P." *European Journal of Epidemiology,* vol. 31, no. 4, 2016, pp. 337-350.

Greenland, Sander, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman. "Statistical Tests, p Values, Confidence Intervals, and Power: A Guide to Misinterpretations." *European Journal of Epidemiology*, vol. 31, no. 4, 2016, pp. 337–350.

Gubits, Daniel, Winston Lin, Stephen Bell, and David Judkins. "BOND Implementation and Evaluation: First and Second-Year Snapshot of Earnings and Benefit Impacts for Stage 2." Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, 2014.

Gubits, Daniel, Rachel Cook, Stephen Bell, Michelle Derr, Jillian Berk, Ann Person, David Stapleton, Denise Hoffman, and David Wittenburg. "BOND Implementation and Evaluation: Stage 2 Early Assessment Report." Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, 2013.

Hock, Heinrich, David Wittenburg, Kate D'Anello, Ken Fortson, Laura Kosar, Martha Kovac, Michael Levere, David Stapleton, and Heather Gordon. "Promoting Opportunity Demonstration Task 2.1. Recruitment and Enrollment Plan." Princeton, NJ: Mathematica Policy Research, 2017.

Hoffman, Denise, Sarah Croake, David R. Mann, David Stapleton, Priyanka Anand, Chris Jones, Judy Geyer, Daniel Gubits, Stephen Bell, Andrew McGuirk, David Wittenburg, Debra Wright, Amang Sukasih, David Judkins, and Michael Sinclair. "2016 Stage 1 Interim Process, Participation, and Impact Report." Washington, DC: Mathematica Policy Research, 2017.

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

Isaacs, Julie. "The Costs of Benefit Delivery in the Food Stamp Program: Lessons from a Cross-Program Analysis." Washington DC: Economic Research Service, U.S. Department of Agriculture, 2008.

Livermore, Gina, Debra Wright, Allison Roche, and Eric Grau. "2006 National Beneficiary Survey: Methodology and Descriptive Statistics." Report no. 4. In *Work Activity and Use of Employment Supports under the Original Ticket to Work Regulations*. Washington, DC: Mathematica Policy Research, 2009.

McConnell, Sheena, Elizabeth Stuart, Kenneth Fortson, Paul Decker, Irma Perez-Johnson, Barbara Harris, and Jeffrey Salzman. "Managing Customers' Training Choices: Findings from the Individual Training Account Experiment." Washington, DC: Mathematica Policy Research, 2006.

Michalopoulos, Charles, David Wittenburg, Dina Israel, Jennifer Schore, Anne Warren, Aparajita Zutshi, Stephen Freedman, and Lisa Schwartz. "The Accelerated Benefits Demonstration and Evaluation Project: Impacts on Health and Employment at Twelve Months." New York: MDRC, February 2011.

Raudenbush, Stephen W., and Howard S. Bloom. "Learning About and From a Distribution of Program Impacts Using Multisite Trials." *American Journal of Evaluation*, vol. 36, no. 4, 2015, pp. 475-499.

Robins, James, Mariela Sued, Quanhong Lei-Gomez, and Andrea Rotnitzky. "Comment: Performance of Double-Robust Estimators When 'Inverse Probability' Weights Are Highly Variable." *Statistical Science,* vol. 22, no. 4, November 2007, pp. 544–559.

Rubin, Donald B. "The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies," *Biometrics*, vol. 29, No. 1, March 1973, pp. 185-203.

Särndal, Carl-Erik, Bengt Swensson, and Jan Wretman. Model-Assisted Survey Sampling. New York: Springer-Verlag, 1992.

Stapleton, David C., Bonnie L. O'Day, Gina A. Livermore, and Andrew J. Imparato. "Dismantling the Poverty Trap: Disability Policy for the 21st Century." The Milbank Quarterly, vol. 84, no. 4, 2006, pp. 701-732.

Storey, John. "The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value." *The Annals of Statistics*, vol. 31, no. 6, 2003, pp. 2013-2035.

Wasserstein, Ronald L., and Nicole A. Lazar. "The ASA's Statement on p-Values: Context, Process, and Purpose." *The American Statistician*, vol. 70, no. 2, 2016, pp. 129-133.

Weathers Robert and Hemmeter, Jeffrey. "The impact of changing financial work incentives on the earnings of social security disability insurance (SSDI) beneficiaries." *Journal of Policy Analysis and Management*, vol. 30, no. 4, 2011, pp. 708–728.

Wiseman, Michael. "Rethinking the Promoting Opportunity Demonstration Project." Washington, DC: Social Security Advisory Board, 2016.

Wittenburg, David, David R. Mann, and David Stapleton. "Work Incentive Simplification Pilot: Recommendations of the Technical Advisory Panel Regarding the Evaluation Design." Princeton, NJ: Mathematica Policy Research, April 25, 2012.

Wittenburg, David, David R. Mann, David Stapleton, Daniel Gubits, David Judkins, and Andrew McGuirk. "BOND Implementation and Evaluation: Third-Year Snapshot of Earnings and Benefit Impacts for Stage 1." Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, 2015.

Wittenburg, David, Jeffery Hemmeter, Holly Matulewicz, Lindsay Glassman, and Lisa Schwartz. "When Every Dollar Counts: Comparing Reported Earnings of Social Security Disability Program Beneficiaries in Survey and Administrative Records." Washington, DC. Mathematica Policy Research, July 2016.

Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2010.

Wright, Debra, Gina Livermore, Denise Hoffman, Eric Grau, and Maura Bardos. "2010 National Beneficiary Survey: Methodology and Descriptive Statistics Final Report." Washington, DC: Mathematica Policy Research, April 2012.

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

http://www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

APPENDIX A

SUPPLEMENT TO CHAPTER II

This page has been left blank for double sided copying.

In this appendix we summarize the theoretical predictions of the POD offset on outcomes, including potential differences in outcomes for key subgroups summarized in Chapter II. We develop theoretical predictions of the effect of the new POD offset on outcomes based on a neoclassical economic model that compares the (current law) cash cliff for the control group with the new POD offset ramp for the two treatment groups.

We first focus on the predicted effects of the POD offset rules for those beneficiaries who are most likely to benefit from POD, whom we define as those beneficiaries who are facing the cash cliff under current rules (that is, those who completed the TWP and Grace Period and are able to engage in SGA). This group is a natural starting point because these beneficiaries have a strong incentive to participate in POD given the POD offset includes a new benefit adjustment process that allows some beneficiaries to keep more benefits while working and makes other changes to current rules (e.g., eliminating the TWP).

We then consider other theoretical assumptions to show how other beneficiary subgroups might respond under POD relative to those in current rules. For example, those who are still within the TWP would always be better off under current rules while in the TWP than under POD. We illustrate examples of different scenarios to show changes in incentives. As noted in Chapter III, the BOND experience indicates that a mix of potential beneficiaries might join POD, including those still in the TWP. Consequently, beneficiary responses could vary from the economic model presented for a simple, post-TWP example.

We conclude with a summary of predicted outcomes, which matches the predictions shown in Chapter II. Because of the complexity of the current rules and the heterogeneity of characteristics of the beneficiary population, particularly in regards to completing the TWP (or expectations around completing the TWP), predicted signs for impacts on many outcomes are ambiguous.

A. Neoclassical economic model with a POD volunteer facing the cash cliff under current rules

As a starting point, we show the economic incentives using a neoclassical model of the POD offset compared with current rules for a beneficiary who would be facing a cash cliff under current rules. The neoclassical model shows a labor–leisure trade-off. In this trade-off, every person has a wage, *w*. The person chooses how to divide his or her time between hours of paid work and hours not at work, termed "leisure" for simplicity, but encompassing all unpaid activities.

Exhibit A.1 shows beneficiary budget constraints—how a beneficiary's income depends on the number of hours the beneficiary works—under both current law and the POD offset. The exhibit illustrates the type of beneficiary likely to benefit from the POD offset, and therefore likely to volunteer for POD. In particular, we focus on an example of a beneficiary who is not blind; is not eligible for SSI; faces the cash cliff (that is, completed the TWP and Grace Period); has no Impairment-Related Work Expenses affecting countable earnings; and is capable of working enough hours to make the POD offset more desirable relative to current law. The budget constraints and indifference curves will vary among these potential volunteers. We start with an example exhibiting the possible positive impacts of the POD offset on earnings and employment

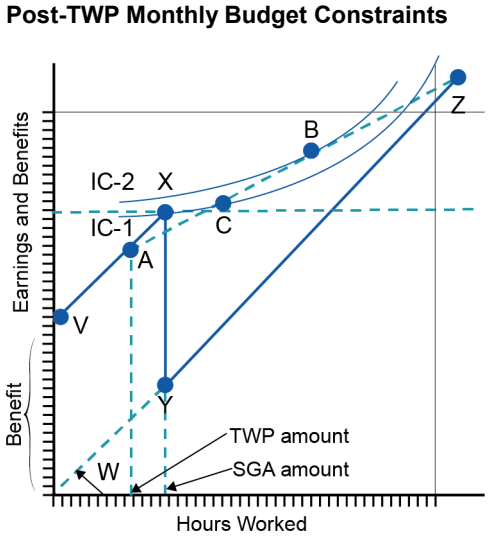
outcomes. Because POD is voluntary, we expect beneficiaries that fall into the categories above will likely volunteer at higher rates than other volunteers, which is an assumption we can directly test in the participation analysis.

We initially simplify other aspects of POD rules so that we can narrow in on predictions for impacts of the POD offset among volunteers. Specifically, we hold constant the main potential effect of the eligibility termination conditions that apply to the second POD treatment group, but not the first. In addition, we hold constant several other factors that might affect impacts. These include the fixed costs of work and the so-called lumpiness of job opportunities; the effects of current work on future earnings; improvements in the functioning of the administrative process for adjusting benefits, primarily due to eliminating the TWP and Grace Period; and taxes.

As a starting point, we compare income differences based on earnings under current rules and the POD offset. We define total income as the sum of SSDI benefits and earnings on the y-axis. To simplify the exposition, we assume the wage rate *w* equals 1; that is, earnings increases 1 unit for a 1 unit increase in work. If a beneficiary is not working (and thus has no earnings), the beneficiary receives his or her full SSDI benefit—point V on the vertical axis. Under current law, income rises with earnings at a $1 for $1 rate until the beneficiary reaches the cash cliff. At low levels of hours worked, the SSDI benefit is unchanged. In this range, total income is the sum of earnings and the full SSDI benefit, and total income increases by *w* ($1, in this simplified example) for each hour worked (from points V to point X). Once earnings exceed the cash cliff, the SSDI benefit under current law drops to zero and total income drops to earnings alone (from point X to point Y). The cash cliff begins at the SGA amount after the duration of the Grace Period. For earnings above the SGA amount, total income is equal to earnings—the solid diagonal line from the right of point Y, along which income again increases with earnings at a $1 for $1 rate.

Under the POD offset, income also continues to rise with earnings at $1 for $1 rate until a person earns up to the TWP amount, but changes after the TWP (POD threshold). The implication is that the current law and POD offset overlap from point V to point A. After the POD threshold, income continuously rises as hours increase beyond point A (where earnings are equal to the TWP amount), past the benefit cliff at point X and up to point Z. This is represented by the dashed line, constituting the POD offset's budget constraint over this range of hours worked. In this range, income increases by $1 for every $2 in additional earnings, as the benefit offset reduces benefits by $1 for every $2 in earnings above the TWP amount until hours reach the level corresponding to full offset, which is point Z. Thus, the POD offset eliminates the cash cliff.

Exhibit A.1. The post-TWP budget constraints and predicted choices of hypothetical non-blind SSDI-only beneficiary under current law and the POD benefit offset



We have added indifference curves to show beneficiaries' possible responses to current law and the POD offset. Each point on the indifference curve depicts the combinations of hours worked/income that are equally desirable for a hypothetical beneficiary. We intentionally set the first indifference curve (IC-1) to cross the SGA earnings threshold, point X, to help show a hypothetical beneficiary's possible response under current law and the new offset above and below the SGA earnings threshold.

The budget constraint under current law creates a strong disincentive to work hours if the corresponding earnings are only modestly larger than the SGA because of the cash cliff, as illustrated by IC-1. In this model, the beneficiary prefers points above and to the left of IC-1 because he or she prefers more income and fewer hours of work. This hypothetical beneficiary is better off at point X than at any other point on the budget constraint under current law. The preferences of this beneficiary are such that, under current law, he or she would not choose to earn more than the SGA amount. Neoclassical theory allows for beneficiaries who are willing to give up their benefits for work under current law; for such a beneficiary, the indifference curves would be flatter, indicating a greater willingness to trade off leisure for higher income.

The POD offset creates new incentives for the hypothetical beneficiary shown in Exhibit A.1 to earn above the SGA amount (at point X), along the dashed portion of the POD budget constraint. We show this point by adding a new indifference curve, IC-2. IC-2 is to the left of IC-1, with higher income for any given level of hours worked. This implies that the beneficiary prefers all points on IC-2 to IC-1. In other words, any point on IC-2 makes the beneficiary better off relative to IC-1.

In summary, the beneficiary depicted in the graph is always better off under POD given the move to a higher indifference curve, which results in positive employment increases and reductions in benefits. Specifically, because this hypothetical beneficiary can now choose hours corresponding to point B on IC-2, he or she would choose to do so under the POD offset.

Compared with the beneficiary's choice of point X under current law, under the POD offset, the beneficiary attains a preferred combination of leisure and income, works more hours, earns more, has lower benefits, and has higher income (that is, the sum of earnings and benefits).

B. Other theoretical considerations

In this section we apply the theoretical model described above to consider examples of beneficiaries with different profiles, including those for whom determining benefits requires more complex information and calculations. The neoclassical model implies that it is possible to identify subgroups of beneficiaries who would not benefit from the POD offset if POD were a mandatory national policy. These groups are important to consider because of the negative implication of the POD offset for their economic well-being if POD rules (i.e., the POD offset and other POD changes, such as the elimination of the TWP) were mandatory. Understanding how the POD offset affects such groups is important because of the implications for interpreting the findings for the evaluation. For example, because POD is voluntary, the number of beneficiaries in these groups who willingly participate in POD is likely to be small relative to their representation in the national population. However, some will likely volunteer, because at the point of POD enrollment they might be optimistic that the POD offset provides them opportunities that are more desirable than those available under current law. Further, if they do volunteer and are assigned to a treatment group, they may revert to current law when they discover that no opportunities under the POD offset are better than those under current law. For symmetric reasons, some beneficiaries who would prefer some opportunities available under the POD offset to all those available under current law might not volunteer for POD.

In this section we also discuss how the POD rules, which includes the POD offset and other POD changes (see Chapter II), could affect behavior in ways that differ from the predictions of the basic neoclassical model. In particular, simplifying the rules could have an effect on employment and earnings behavior that is independent of the financial incentives that underpin the graphical example in the previous section. For example, the experience of BOND volunteers shows that these alternatives are important. Some volunteers in BOND never completed their TWP, though the expectation for BOND, as for POD, was that the volunteers would largely consist of those beneficiaries most likely to benefit from the new earnings rules. Hence, it is important to consider that people might volunteer for POD for reasons other than those of the hypothetical beneficiary above and complicate predictions for the overall beneficiary groups.

Predicted impacts for beneficiaries with different wage rates, benefits levels, or preferences. The predictions associated with Exhibit A.1 depend on the specific indifference curves and budget constraints for individual beneficiaries. Beneficiaries who have sufficiently lower wage rates, benefits, or willingness to give up leisure in exchange for income than the depicted hypothetical beneficiary might find that the POD offset does not provide better opportunities than current law and might be less likely to volunteer. Changing any one of these features graphically by a sufficient amount for the hypothetical beneficiary would result in IC-1 lying entirely above the POD budget constraint. As we will discuss in more detail below, the potential variation in indifference curves based on beneficiary circumstances is important for theoretical predictions.

Earnings below TWP amount. The neoclassical model has implications for the large percentage of beneficiaries whose hours worked are less than the hours corresponding to their TWP amount, *including the majority who do not work at all*. Given their choice under current law, the model implies that the amount they would earn for an hour of work (the slope of their budget constraint at every point except X) is less than the minimum they would be willing to accept for the first hour of work. The latter amount is called their reservation wage, which is the slope of the indifference curve passing through point V (zero hours and earnings) on their budget constraint combined with the neoclassical properties of indifference curves. In other words, based on this model we should not expect more beneficiaries to work under POD rules than do under current law. Following similar reasoning, the model predicts that those who would work under current law but never earn as much as the TWP amount would behave no differently under the POD offset.

Earnings between TWP and SGA amounts. Another feature that distinguishes the POD budget constraint from the current-law budget constraint is that it includes a set of points between TWP hours and SGA hours that are *below* the current-law budget constraint. Holding earnings constant, total income under the POD design is less than it is under current law for any given hours worked within this range. If the POD design were to replace the current-law design for all beneficiaries, the model implies that some beneficiaries who would choose hours worked in this range under current law would be worse off under the POD design. Relative to the depicted hypothetical beneficiary, the wages, benefits, or willingness to enter work in exchange for income for such beneficiaries are such that these beneficiaries would prefer no points on the POD budget constraint with hours worked above SGA hours over the combination of work hours and income they would choose under current law (between points A and X on the current-law budget constraint).

Earnings above SGA. Finally, the neoclassical model predicts that many of those who work enough hours under current law to experience benefit suspension or, eventually, termination will receive a partial benefit under POD, even if they continue to work and earn the same amount. Beneficiaries who would choose a point on their current-law budget constraint between points Y and Z would receive a partial benefit with the POD offset if they work and earn exactly the same amount. The model also predicts that such beneficiaries are likely to reduce their hours and earn less under the POD offset, for two reasons: (1) the increase in their benefit reduces the value of an additional dollar of income, and (2) when their earnings drop by a given amount, their income drops by only half as much as it would under current law. The latter effect also applies to those who would earn just above the point represented by Z under current law. We expect some beneficiaries who would work hours to the right of point Z and thus not receive any benefits under current law would instead reduce their hours under the POD offset enough that they receive a partial SSDI benefit.

Other characteristics affecting predicted impacts. Other beneficiary characteristics are likely to affect impacts for some volunteers, but the same characteristics may mean that few such beneficiaries will volunteer. For example, the treatment of Impairment-Related Work Expenses under the POD design is likely to reduce the likelihood of volunteering among those with high Impairment-Related Work Expenses, other things constant, and could affect how those who do volunteer respond to the POD design (see Chapter III for more details). Similarly, because blind beneficiaries have higher SGA amounts, they are less likely to volunteer, other things constant,

and the behavioral responses of those who do volunteer could differ because of the higher SGA amount (see predictions above for those below SGA).

Predicted impacts of POD termination provisions. A feature of POD that is difficult to show in the neoclassical model is the elimination of the SSDI eligibility termination due to work for the first treatment group. Specifically, this feature of POD could further reduce the uncertainty that beneficiaries face in making work decisions. For example, if POD changes beneficiary perceptions about loss of benefits—even if that perception is incorrect under current law for those in the TWP—POD could lead to employment increases beyond those described above.

Between treatment groups, mean earnings and income will be lower and mean benefit payments higher under the POD offset with termination conditions than they would under the POD offset without termination conditions. This is primarily because some beneficiaries might not want to go through the process of re-entering SSDI if their benefits are terminated for work. More specifically, we predict that, if the termination conditions apply: (1) there will be fewer 12-month periods with no benefits due to earnings; (2) the percentage of beneficiaries earning at least *P* percent of the smallest earnings amount that results in no benefit payment will be no larger than the corresponding percentage if the termination conditions do not apply; and (3) that any difference in *P* across groups will increase in magnitude as *P* approaches 100 percent. We also note that the expedited reinstatement provisions (including provisional benefit payments) that apply for 60 months after termination for work, as under current law, reduce the risk of termination.

C. Summary of predicted effects on primary outcomes

In summary, the predictions for certain subgroups of beneficiaries have clear theoretical predictions, particularly those who face the cash cliff under current rules. Holding all else equal, the theory predicts higher rates of volunteering for POD and more positive earnings impacts for beneficiaries who have completed the TWP and Grace Period, have higher wage rates, have higher monthly benefit amounts, have few or no Impairment-Related Work Expenses, and are not blind.

However, similar to BOND, the predicted signs of impacts for many mean outcomes are ambiguous for the overall population and will depend on the extent to which volunteers comprise beneficiaries from the subgroups most likely to have better economic opportunities under the POD offset. Impacts on earnings are likely to be positive if volunteers predominantly consist of such beneficiaries. Whether or not the earnings impacts for volunteers are positive, they are likely to be more positive than they would be for the full population of SSDI beneficiaries under a mandatory benefit. This is because beneficiaries for whom impacts on earnings are likely to be zero or negative are less likely than others to volunteer.

APPENDIX B

SUPPLEMENT TO CHAPTER III

This page has been left blank for double sided copying.

Exhibit B.1. Catchment areas for Alabama

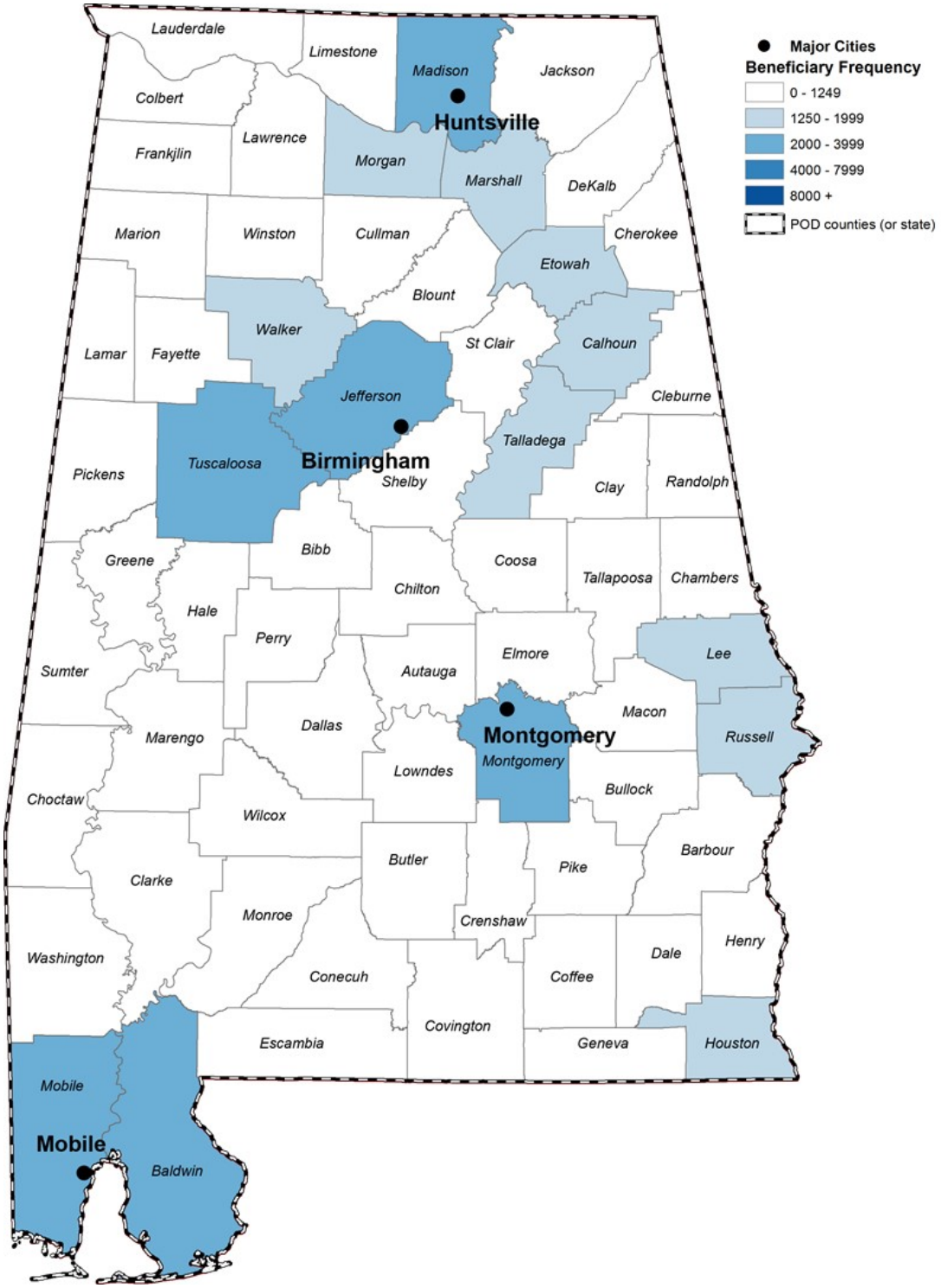


Exhibit B.2. Catchment areas for California

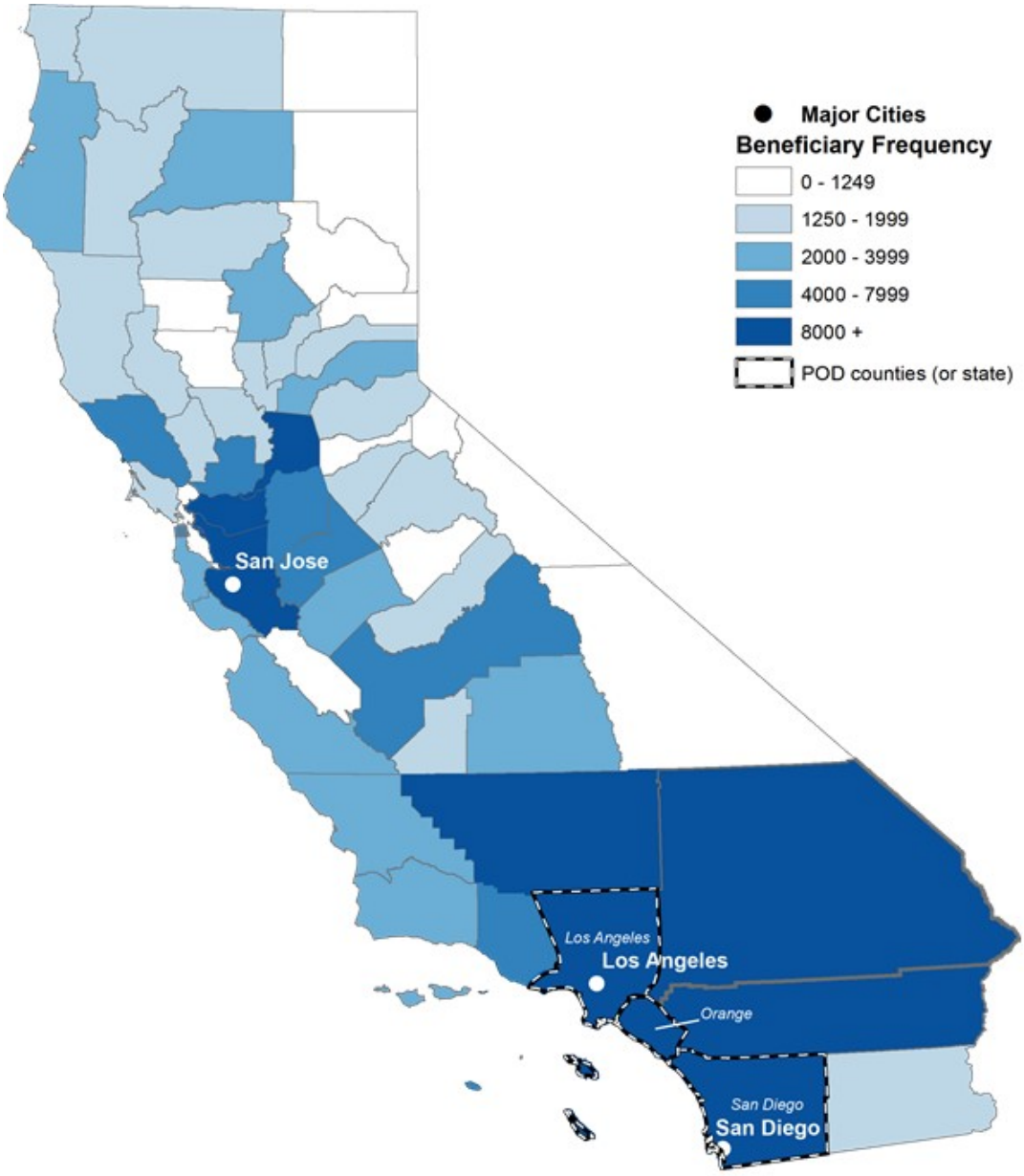


Exhibit B.3. Catchment areas for Connecticut

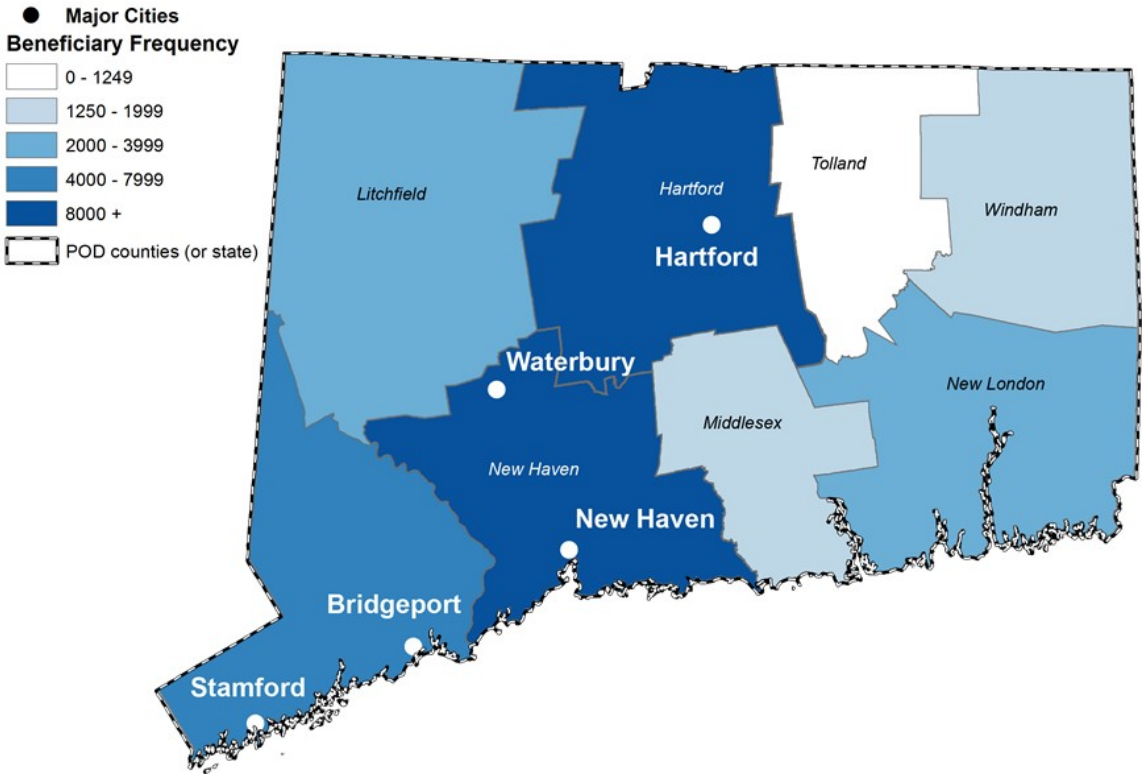


Exhibit B.4. Catchment areas for Maryland

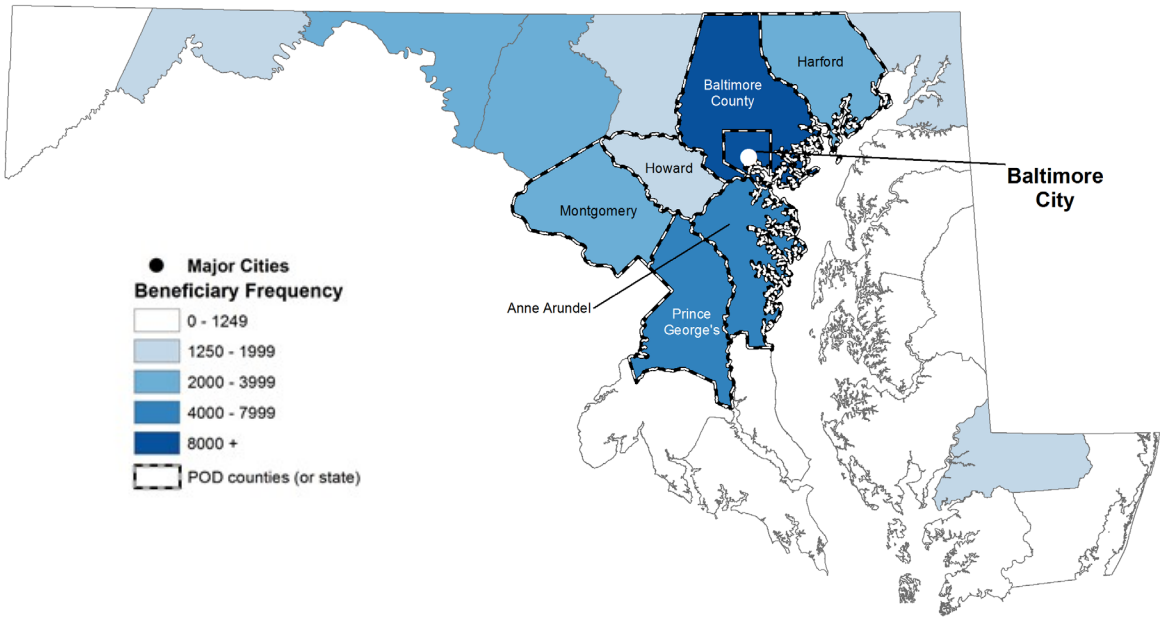


Exhibit B.5. Catchment areas for Michigan

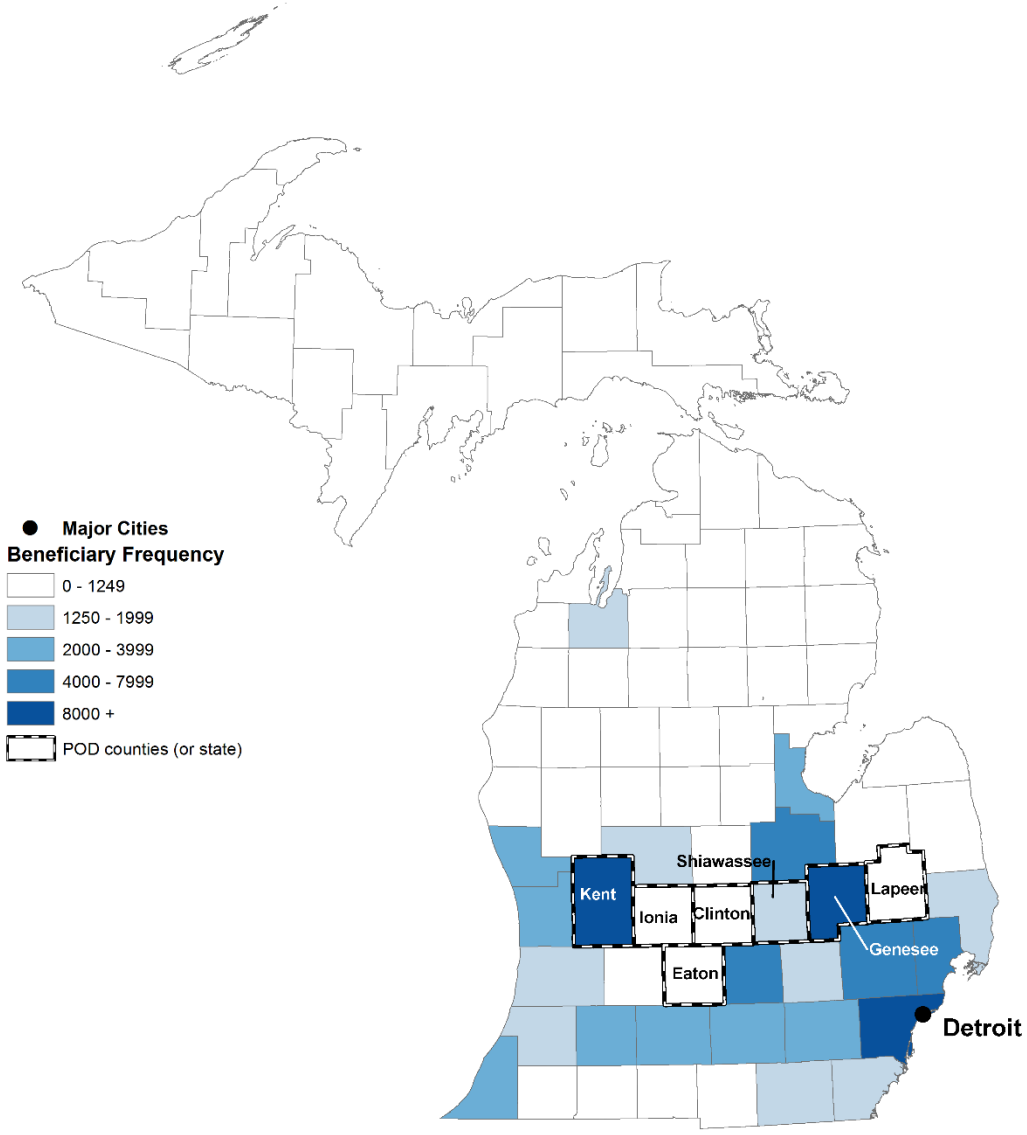


Exhibit B.6. Catchment areas for Nebraska

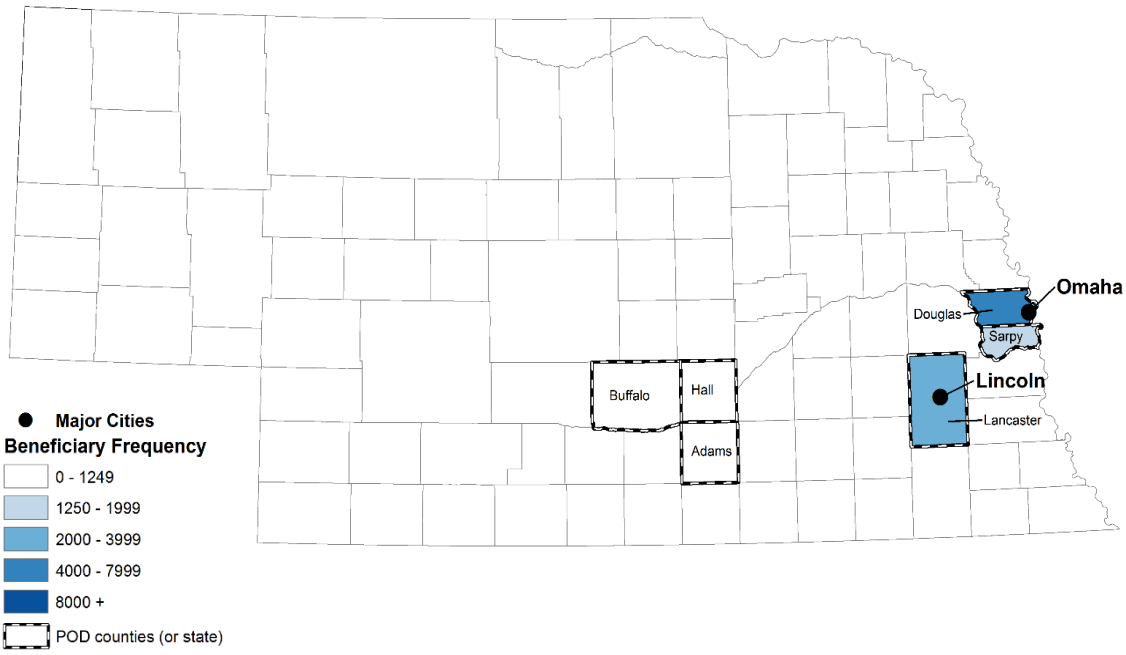


Exhibit B.7. Catchment areas for Texas

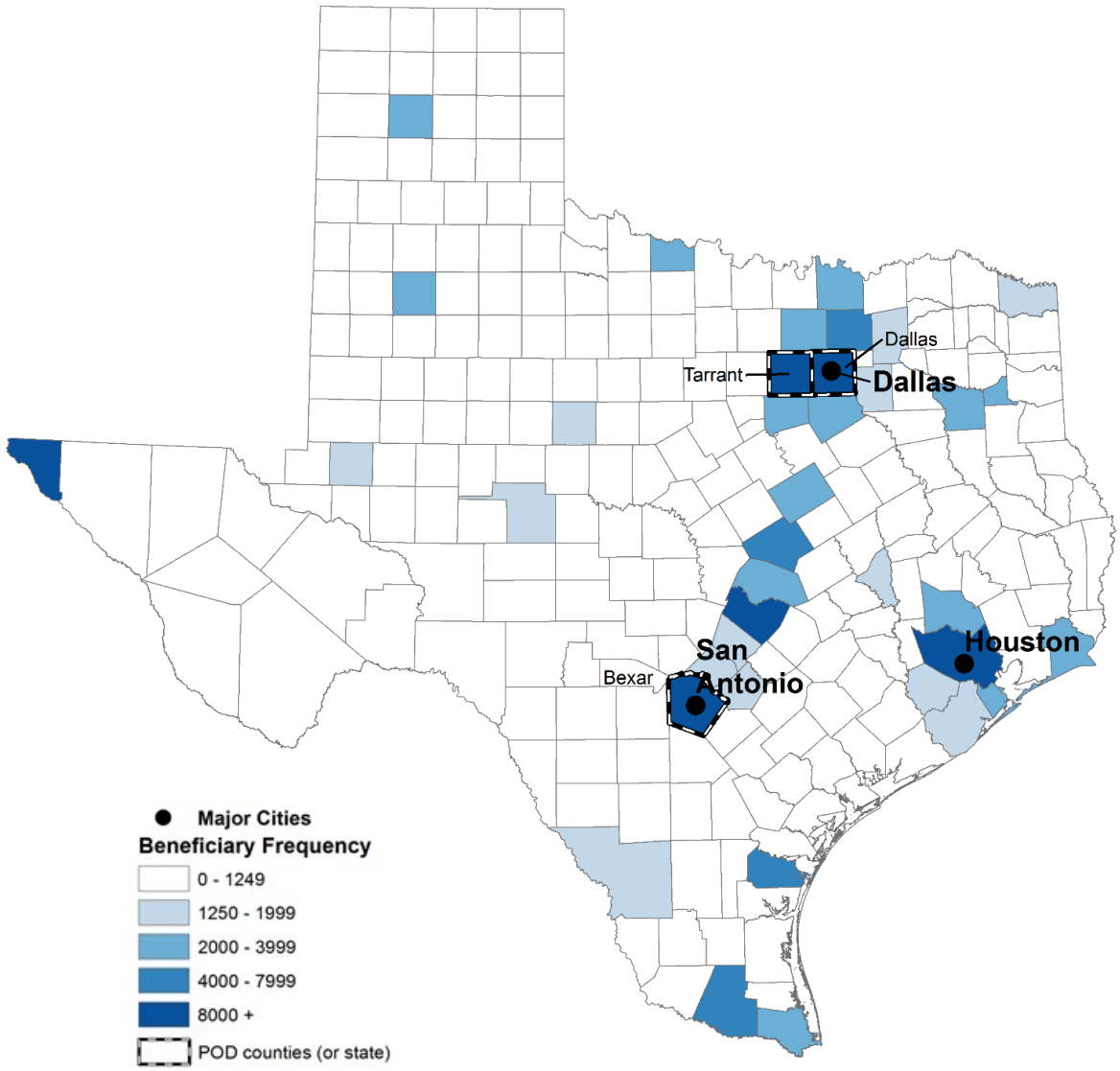
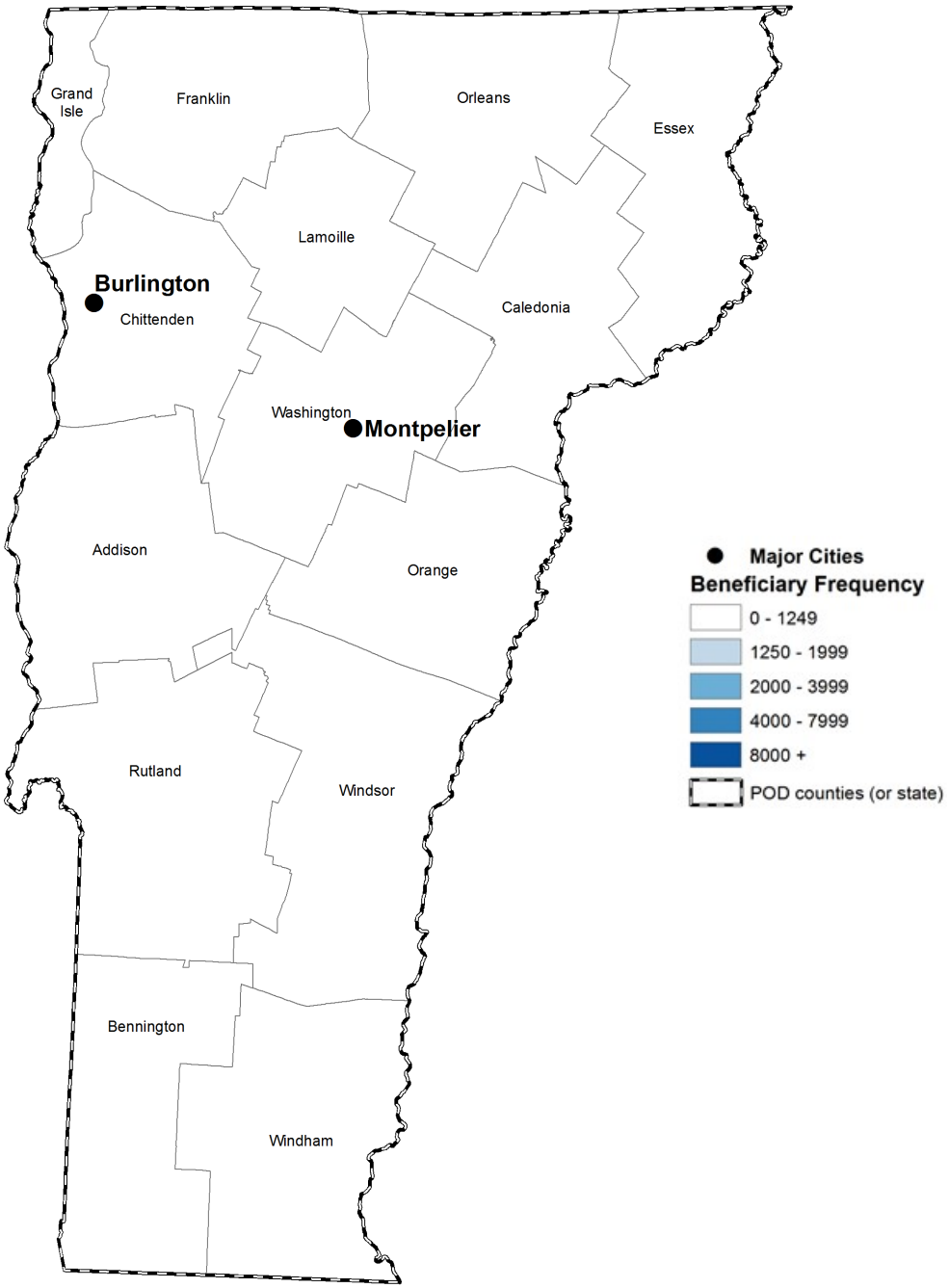


Exhibit B.8. Catchment areas for Vermont



APPENDIX C

SUPPLEMENT TO CHAPTER V

This page has been left blank for double sided copying.

In this appendix we summarize our plans for conducting site visits. Our proposed site visits will require detailed planning and effective coordination with demonstration partners in each of the POD states (Exhibit C.1). Approximately three months before the first round of site visits in early 2018, the state liaisons will participate in a conference call with the Virginia Commonwealth University site director and VR agency/WIPA manager in each POD state to discuss Mathematica's data collection plans. Shortly after the call, the designated state liaison will send an email to the state VR agency/WIPA provider point of contact for each POD state. The email will describe site visit activities, identify the approximate timeframe for the visit, and request a date for a planning meeting via telephone to discuss the logistics of the site visit and all site visit activities. During the planning meeting with the state VR agency/WIPA provider point of contact, we will discuss the schedule for the visit (for example, length of interviews with each key informant and each informant's role and responsibilities within the organizational structure of the state VR agency/WIPA provider) and learn where each key informant is located within the catchment area. We will also inquire if there are other key stakeholders, such as representatives from the local American Job Center, Centers for Independent Living, or local employment network, who could offer valuable perspectives on the local service context and potentially participate in an interview. After these initial meetings, the state liaisons will follow-up by email and telephone to coordinate logistics for the site visits.

Exhibit C.1. Site visit planning activities

Weeks before site visit	Scheduling activity	Purpose of activity	Demonstration partners involved
12	Participate in a conference call with the POD site director and VR agency/WIPA points of contact	• Provide overview of evaluation objectives and site visit data collection plans	• POD site director • Virginia Commonwealth University site liaison • VR agency/WIPA point of contact • Mathematica state liaison
11	Send follow-up email to state VR/WIPA point of contacts	• Provide overview of site visit activities • Propose site visit dates • Propose dates/times for planning meeting with VR/WIPA point of contact during week 10	• Mathematica state liaison • VR agency/WIPA point of contact
10	Send advance email to state VR/WIPA point of contact and follow-up by telephone during scheduled meeting time	• Provide overview of site visit activities and respondents to participate in interviews • Learn where respondents are geographically located • Identify local stakeholders (American Job Centers, Centers for Independent Living, Employment Networks) who might offer valuable perspectives of local service environment • Review timeframe for data collection • Request program documents	• Mathematica state liaison • VR agency/WIPA point of contact
3-9	Follow-up communication, as needed	• Planning and preparation for site visit, including making travel arrangements, tailoring interview protocols, and reviewing background materials	• Mathematica state liaison • VR agency/WIPA point of contact
1-2	Follow up by telephone with state VR/WIPA point of contact	• Confirm any information that might have changed • Provide site visitor's name and contact information • Discuss site visit activities and schedule, including staff interviews and observation of site operations (i.e., benefits counseling sessions) • Review site visit logistics one final time	• Mathematica state liaison • VR agency/WIPA point of contact

A. Pilot testing

For the first round of data collection, we will pilot test the interview protocols by conducting a site visit to California in March 2018. Abt suggested California as the pilot site because it had a relatively high number of enrolled treatment subjects and started implementation early in the pilot period.

The pilot site visit has several important objectives including an assessment of: (1) communication and coordination strategies used with the demonstration VR/WIPA director, POD counselors, VR/WIPA manager supervising the POD counselors, Virginia Commonwealth University TA liaisons, and local VR agency and other stakeholders for planning site visit activities, (2) site visitors' ability to collect the information needed in the allotted time, (3) whether respondents can readily understand and answer the interview questions, (4) whether interviews flow sensibly from topic to topic, and (5) whether the questions yield thoughtful, candid responses. The pilot will also be useful for identifying site visitor training needs. We plan to conduct the pilot site visit in February 2018, the last month of the pilot period, to observe site operations immediately before full implementation in March 2018. The timing of the pilot site visit allows us ample time to modify data collection procedures based on our findings prior to the first round of data collection during full enrollment (expected to begin in late March 2018).

B. Site visitor trainings

Customized, comprehensive training is vital for uniform, consistently high-quality data collection (Exhibit C.2). We will conduct five training sessions corresponding to the following four topics: (1) site visit preparation procedures, (2) conducting the site visits, which will be delivered during two separate trainings, (3) the research objectives, focal research questions, and use of the consolidated framework for implementation research (CFIR), and (4) coding and analyzing the qualitative data. The training pertaining to coding and analyzing qualitative data will be attended by three to four staff who will be part of the coding team. The five state liaisons will attend all other training sessions. The content of the training also will be informed by our pilot site visit described above. The training sessions will review the semi-structured interview guides, the observational guide, and the data coding schemes. We will also practice with role-playing interviews and discuss how to respond to unexpected events while on site. The site visit trainings will facilitate each team member sharing a common understanding of the goals of the site visits and what is expected of them as researchers/site visitors.

C. Site visit summaries

State liaisons will prepare a site visit summary and submit it to SSA within two weeks after each site visit. The summary will follow a standardized template, and will include counts of T1, T2, and C subjects in each site; a summary of Work Incentives Counseling and offset use among T1 and T2 subjects in each site; the local employment, service, and program environment; the organizational structure and staffing configuration in each VR agency/WIPA provider; the processes and procedures that are implemented to support POD; perspectives on facilitators and barriers to implementation; and views on early demonstration outcomes such as POD offset use and delivery of work incentives counseling. The process study task leader of the POD evaluation team will review each site summary to check for internal consistency and completeness of information.

Exhibit C.2. Site visit training topics

Training	Training topics
1. Preparing for site visits	• Background on POD • Conducting planning call with demonstration sites • Preparation for site visits, including booking travel, tailoring protocols, recording and transmitting qualitative data
2. Conducting site visits (Delivered in 2 parts)	• Overview of demonstration partners implementing POD • Background on respondents, roles, and responsibilities • Review of interview protocols • Review of site visit summary template • Schedule and process for preparing site visit summaries • Overview of SSA security requirements and procedures to follow when collecting and transmitting qualitative data • Discussion of safeguards to maintain firewall between implementation and evaluation teams
3. Use of consolidated framework for implementation research (CFIR)	• Introduction to CFIR • How to use CFIR • Review of CFIR domains and constructs • How CFIR is being implemented on POD
4. Coding of qualitative data	• Overview of coding schemes • Review of POD logic model • Review of process for coding qualitative data • Review of process for checking coded notes for inter-rater reliability

APPENDIX D

SUPPLEMENT TO CHAPTER VI

This page has been left blank for double sided copying.

Exhibit D.1. Research Question Codebook

Research Question code	Definition and Coding Rules
1. Pre-POD employment services (provided to SSDI beneficiaries)	Code discussion of employment services provided to SSDI beneficiaries prior to the demonstration, [*date*]. This can also include discussion of general employment and service environment in the local, state policies, and other state-specific contextual features (i.e., public transit) that could influence benefit offset use during the demonstration; Double code with a program component code. **Coding rule:** if respondent includes in their description of the local/service environment external factors that influence the implementation of POD Work Incentives Counseling or the POD benefit offset, code that to the appropriate Contextual Factors code.
2. Other services	Code discussion of other government programs, income supports, or services used by POD participants (before, during, or after enrollment in POD); Double code with a program component code.
3. Intervention description	Code neutral description of the POD intervention including delivery of POD Work Incentives Counseling services, what is being implemented in state VR agencies/WIPA provider organizations; Double code discussion of Work Incentives Counseling services with a Work Incentives Counseling program components code; Double code discussion of POD benefit offset with a POD benefit offset program components code. **Coding rule:** if respondent includes in their description of POD challenges or facilitators to implementing POD Work Incentives Counseling or the POD benefit offset, or a factor that influences the implementation of POD Work Incentives Counseling or the POD benefit offset, code that to the appropriate contextual factors code.
4. POD Infrastructure description	Code description of the POD infrastructure including the components involved in administering or supplying the intervention. Include descriptions of the intake processes to recruit, enroll, and randomize volunteers; the data systems in place to support enrollment of volunteers, service delivery, and site operations; and processes/procedures that support administration of the intervention, including initial training activities and materials. • Double code with a program component code.
5. Need for modification	Code discussion of need to change or modify program implementation. • Double code discussion of changes to Work Incentives Counseling services or service delivery with a Work Incentives Counseling program components code; changes to the POD benefit offset with a POD benefit offset program components code; or changes to administration of the POD infrastructure with a Pre-Intervention activities or POD data systems program components code. **Coding rule:** if respondent includes in their description of POD challenges or facilitators to implementing POD Work Incentives Counseling or the POD benefit offset, or a factor that influences the implementation of POD Work Incentives Counseling or the POD benefit offset, code that to the appropriate contextual factors code.
6. Fidelity	Code discussions that relate to whether or not the program is being implemented as planned, including efforts to maintain integrity of random assignment and extent to which POD counseling services are being delivered consistently across some/all sites; Double code with a program component code.
7. Subjects' view of POD	Code discussion of beneficiaries' views on POD; Double code with a Work Incentives Counseling or POD benefit offset program component code.
8. Subject motivation to enroll in POD	Code discussion of the reasons why beneficiaries agreed to enroll in POD; Double code with a Pre-Intervention activities program components code.
9. Subject motivation to withdraw from POD	Code discussion of the reasons why treatment subjects withdrew from POD; Double code with a program component code.

Note: The coding scheme is subject to change as data collection plans are refined further.

Exhibit D.2. Program Component Codebook

Operational code	Definition and Coding Rules
Pre-intervention activities	
1. POD recruitment strategies	Code discussion of strategies that were used to inform SSDI beneficiaries and local stakeholders about POD. Include discussion of four recruitment experiments the evaluation team is testing under POD.
2. POD enrollment and random assignment processes	Code discussion of processes and data systems supporting enrollment and randomization of participants. Include discussion of baseline survey and informed consent.
Work incentives counseling	
3. Onboarding new subjects (intake procedures)	Code discussion of POD counselors using the (standardized) intake protocol to onboard new subjects and collect necessary information.
4. Collecting and reporting monthly earnings to SSA (and coordinating earnings and impairment-related work expenses	Code discussion of POD counselors working with treatment subjects on their monthly earnings and Impairment-Related Work Expenses reporting and reconciling the earnings and Impairment-Related Work Expenses information that is reported.
5. Assisting treatment subjects with requests for appeals/waivers	Code discussion of POD counselors assisting treatment subjects with appeals or waivers in response to changes made to their benefits payments.
6. Developing benefits summary and analyses (benefits planning query analysis and benefits verification)	Code discussion of POD counselors summarizing case-specific information about the past, current (at on-boarding), and future use of work incentives that support a treatment subject's work and earnings goal, including: • Confirming and summarizing all the federal and state benefits the subject is receiving that could be affected by paid employment. This might include assisting subjects to obtain Benefits Planning Queries and other benefits verification • Confirming and summarizing the subject's current employment or earnings goal(s) • Assessing and documenting how the subject's specific earnings goal(s) will affect all benefits the subject receives • Identifying the specific work incentives applicable to the subject • Recommending employment services that could help the subject achieve their specific earnings goal(s) • Options for resolving any benefit issues • Include discussion of initial benefits summary and analyses and revisions to benefits summary and analyses.
7. Developing work incentives plans	Code discussion of the POD counselor using the recommendations and choices in the benefits summary and analyses to collaborate with the treatment subject to develop an individualized plan, or "to-do" list, for using work incentives to further the subject's financial independence goals. Include discussion of initial work incentives plan and revisions to work incentives plans.

EXHIBIT D.2. (CONTINUED)

Operational code	Definition and Coding Rules
8. Delivering ongoing POD work incentives counseling	Code discussion of POD counselors providing ongoing work incentives counseling to treatment subjects, including: • Providing information and explanation of the POD offset and rules to subjects • Helping subjects make choices about their employment and earnings that meet their career and self-sufficiency goals (in their work incentives plan). • Assisting subjects to access services and supports necessary to achieve their self-sufficiency goals • Referring subjects to employment services. • Referring subjects to employment supports. Helping subjects to understand the effect of various employment choices on their benefit eligibility and payment amount (related to benefits summary and analyses) • Supporting subjects when earnings totals vary month to month therefore changing the offset frequently, and resulting in benefits over- or underpayments. (This might overlap with assisting participants (POD subjects) with requests for appeals/waivers) • Counseling on state-specific benefits, such as state supplements and Medicaid for working subjects • Counseling on Expedited Reinstatement, available to subjects whose entitlement is terminated due to earnings (explaining the Expedited Reinstatement process and assisting them to apply for Expedited reinstatement).
9. Supporting diverse and underserved populations	Code discussion of POD counselors delivering culturally sensitive services to treatment subjects with the following content: • Diverse racial, ethnic, and gender background: Focused on treatment subjects from diverse racial heritage, and various ethnicities. • Diverse linguistic backgrounds: Developed language appropriate to English language learners, non-English speakers, or other beneficiaries who may need translation services or accommodations. • Specific disability categories: Taking into account the unique needs of beneficiaries with various types of disabilities. • Specific accommodation needs: Addressing the needs for various types of accommodations that will enable participants to access and benefit from all POD work incentives counseling and POD Call Center services, including communications, written materials, and technology, in order to benefit from services. • Underserved groups: Addressing the needs of typically underserved populations, such as transition-aged youth; veterans; Native Americans and other racial, ethnic, disability, or socioeconomically disadvantaged groups; or participants in rural or urban areas.
10. Transitioning treatment subjects out of POD	Code discussion of treatment subjects transitioning out of POD, including: • Using an off-boarding checklist • Providing POD counseling to treatment subjects about returning to current program rules.
POD data systems	
11. Using the MIS	Code discussion of using management information systems to collect POD related data.

EXHIBIT D.2. (CONTINUED)

Operational code	Definition and Coding Rules
POD benefit offset	
12 Monthly reporting of earnings and Impairment-Related Work Expenses	Code discussion of treatment subjects monthly reporting of earnings and Impairment-Related Work Expenses to POD, including: • Identification of subjects with earnings over POD threshold who need to report earnings each month • Instrument (POD Monthly Earnings and Impairment-Related Work Expenses Reporting Form) used to collect monthly earnings and Impairment-Related Work Expenses information from eligible treatment subjects • Modes for submitting earnings and Impairment-Related Work Expenses information (mail, fax, IDS, and in-person) to POD • Timeliness of monthly reporting of earning and Impairment-Related Work Expenses (from beneficiary to Abt and from Abt to SSA) • Processes in place to support collection of monthly earnings information (i.e., quarterly reminder letters, monthly email or text reminders, etc.)
13 Processing earnings and Impairment-Related Work Expenses information	Code discussion of processing of treatment subjects' earnings and Impairment-Related Work Expenses information submitted to POD, including: • Scanning and uploading of earnings and Impairment-Related Work Expenses information to IDS • Creation and quality review of earnings records in IDS • Processes to follow-up with subjects to address identified issues • Submission of earnings records to SSA
14. Adjusting DI benefits under POD offset rules	Code discussion of adjusting DI benefits under POD offset rules, including: • Adjustment in monthly benefit payments under POD offset rules • SSA notices explaining changes in DI benefit payments • Overpayments, underpayments, and incorrect payments resulting from monthly benefit adjustments • Benefit termination under POD rules due to work and earnings. Applicable to T2 subjects only.
15. Annual automated reconciliation	Code discussion of automated reconciliation that SSA runs annually in August for the previous year to identify the correct amount of benefits that should have been paid to each subject under POD offset rules.

Note: The coding scheme is subject to change as data collection plans are refined further.

Exhibit D.3. Contextual Factors Codebook

Contextual Code	Definition and Coding Rules
A. Characteristics of the intervention (Administration of POD offset and provision of POD work incentives counseling)	
1. Evidence strength & quality	Stakeholders' perceptions of the evidence supporting the belief that the intervention will have desired outcomes.
2. Relative advantage	Stakeholders' perception of the advantage of implementing the intervention versus an alternative solution.
3. Adaptability	The degree to which the intervention can be adapted, tailored, refined, or reinvented to meet the needs of the VR agency/WIPA provider, POD call center, or indirect support units (e.g., POD processing center).
4. Trial	The ability to test the intervention on a small scale in the VR agency/WIPA provider, and to be able to reverse course (undo implementation) if warranted.
5. Complexity	Perceived difficulty of implementation, reflected by duration, scope, disruptiveness, complexity, and number of steps required to implement.
6. Presentation of intervention	Perceived excellence in how the intervention is explained and presented to those implementing it. Example: POD counselors' perceptions of procedural manuals, and other documents explaining operational processes.
7. Cost	Costs of the intervention and costs associated with implementing that intervention including investment, supply, and opportunity costs.
B. State-specific contextual features	
1. Participant needs & resources	The extent to which participant lacks resources (e.g., employment support needs, accessibility needs, etc.) and barriers and facilitators to meet those needs.
2. External networks	The degree to which the VR agency/WIPA provider is networked with other external organizations.
3. Peer pressure	Competitive pressure from another organization (e.g., other VR agency, Employment Network, WIPA provider) to implement the intervention.
4. External policy & incentives	A broad construct that includes external strategies to spread interventions including policy and regulations (governmental or other central entity), external mandates, recommendations and guidelines, collaborative.
C. Internal context of state VR agency/WIPA provider	
1. Structural characteristics	Organizational characteristics (i.e. accessibility for people with disabilities), age, maturity, and size of an organization. Also include the type of entity implementing POD (VR agency, WIPA provider, lower tier subcontractor), discussions of the management structure supporting POD implementation, or approach to service delivery.

EXHIBIT D.3. (CONTINUED)

Contextual Code	Definition and Coding Rules
2. Networks & communications	The nature and quality of social networks within an organization and the nature and quality of formal and informal communications among POD staff working within the VR agency/WIPA provider or Abt support units.
3. Culture	Norms, values, and basic assumptions of the VR agency/WIPA provider or Abt support units.
4. Implementation climate	The capacity for change, shared receptivity of involved individuals to an intervention and the extent to which use of that intervention will be rewarded, supported, and expected within their organization (i.e., VR agency/WIPA provider or Abt support units).
a. Tension for change	The degree to which stakeholders perceive the pre-POD WIPA services or SSA work incentives under current DI program rules as needing change.
b. Compatibility	The degree of tangible fit between meaning and values attached to the intervention by involved individuals; how those align with individuals' own norms, values, and perceived risks and needs; and how the intervention fits with existing workflows and systems.
c. Relative priority	Individuals' shared perception of the importance of the implementation within the organization (i.e., VR agency/WIPA provider or Abt support units).
d. Organizational incentives & rewards	Extrinsic incentives such as goal-sharing awards, performance reviews, promotions, and raises in salary and less tangible incentives such as increased stature or respect.
e. Goals & feedback	The degree to which demonstration goals are clearly communicated, acted upon, and fed back to POD counseling and support staff and alignment of that feedback with goals.
f. Learning climate	A climate in which: a) leaders express their own fallibility and need for team members' assistance and input; b) team members feel that they are essential, valued, and knowledgeable partners in the change process; c) individuals feel psychologically safe to try new methods; and d) there is sufficient time and space for reflective thinking and evaluation.
5. Readiness for implementation	Tangible indicators of organizational (i.e., VR agency/WIPA provider) commitment to its decision to implement an intervention.
a. Leadership engagement	Commitment, involvement, and accountability of POD Implementation leaders and managers overseeing implementation.
b. Available resources	The level of resources dedicated for implementation and on-going operations including money, training, staffing, equipment, education, physical space, and time.
c. Access to knowledge and information	Ease of access to information and knowledge about the intervention and how to incorporate it into work tasks.

EXHIBIT D.3. (CONTINUED)

Contextual Code	Definition and Coding Rules
D. Characteristics of individuals implementing the intervention (POD counselors, POD managers, and POD indirect and direct support staff)	
1. Knowledge & beliefs about the intervention	Individuals' attitudes toward and value placed on the intervention as well as familiarity with facts, truths, and principles related to the intervention. Example: [Believe that DI beneficiaries' receipt of accurate and complete information about work supports and work incentives will allow them to make informed decisions about working and increasing their earnings.]
2. Self-efficacy	Individual belief in their own capabilities to execute courses of action to achieve implementation goals.
3. Individual stage of change	Characterization of the phase an individual is in, as he or she progresses toward skilled, enthusiastic, and sustained use of the intervention.
4. Individual identification with the organization	A broad construct related to how individuals perceive the organization (i.e., VR agency/WIPA provider or Abt Associates for the POD support units) and their relationship and degree of commitment with that organization.
5. Other personal attributes	A broad construct to include other personal traits such as tolerance of ambiguity, intellectual ability, motivation, values, competence, capacity, and learning style.
E. Implementation process	
1. Planning	The degree to which a purposeful method and tasks for implementing an intervention are developed in advance and the quality of those methods. Include discussion of activities related to staff training, planning for implementation, and early implementation activities during the pilot period. Do not include training activities that occur after implementation begins.
2. Engaging	Attracting and involving POD implementation staff in the implementation and use of the intervention through a combined strategy of social marketing, education, role modeling, training, and other similar activities.
a. Opinion leaders	Individuals in an organization who have formal or informal influence on the attitudes and beliefs of their colleagues with respect to implementing the intervention.
b. Formally appointed internal implementation leaders	Individuals from within the organization who have been formally appointed with responsibility for implementing an intervention, such as POD counselors, POD managers, or other similar role.
c. Champions	"Individuals who dedicate themselves to supporting, marketing, and 'driving through' an [implementation]" [101](p. 182), overcoming indifference or resistance that the intervention may provoke in an organization.

EXHIBIT D.3. (CONTINUED)

Contextual Code	Definition and Coding Rules
d. External change agents – technical assistance provided by Abt, Virginia Commonwealth University, and SSA	Individuals who are affiliated with an outside entity who formally influence or facilitate intervention decisions in a desirable direction. The Virginia Commonwealth University Site Liaisons will be responsible for monitoring the performance of the POD sites and delivering technical assistance when they identify a need. Code discussion of the technical assistance provided by Abt and Virginia Commonwealth University and site monitoring. SSA and Abt will also provide policy and operational guidance that will alter/influence how the intervention is implemented. Code discussion of policy or operational guidance provided by SSA or Abt Associates.
3. Executing	Carrying out or accomplishing the implementation according to plan.
4. Reflecting & evaluating	Quantitative and qualitative feedback about the progress and quality of implementation accompanied with regular personal and team debriefing about progress and experience.
5. Training and/or unmet training needs	Use the staff-specific codes to capture discussion of POD training or unmet training needs for each type of staff.
a. VR/WIPA manager training	Code discussion of VR managers receiving: • 8 hours of training on the basic POD design and procedures, referred to as POD 101. • POD IDS User Training, including both general information on using IDS and customized training on role-based functionality.
b. POD counselor training	Code discussion of POD counselors receiving: • 8 hours of training on the basic POD design and procedures, referred to as POD 101. • 32-hours of training focused on POD benefits rules to prepare counselors to explain the unique rules in place for POD and the requirements for its two treatment groups • Counselors who are not Certified Work Incentives Counselors attending a comprehensive initial Certified Work Incentives Counselor training and certification course, approximately 200 hours. • POD IDS User Training, including both general information on using IDS and customized training on role-based functionality • NOTE: The Certified Work Incentives Counselor training requirements have been relaxed for POD counselors working in the Maryland POD site. Code discussion of the training that POD counselors in Maryland have received prior to and during implementation of POD.

EXHIBIT D.3. (CONTINUED)

Contextual Code	Definition and Coding Rules
c. Abt call center training	Code discussion of training call center staff receiving: • 8 hours of training on the basic POD design and procedures, referred to as POD 101 • Detailed role-based trainings to prepare them to assist POD callers; the Abt team will provide this training • POD IDS User Training, including both general information on using IDS and customized training on role-based functionality
d. Mathematica toll-free specialist training	Code discussion of training Mathematica toll-free specialists receiving, provided by Mathematica operations staff.
e. POD support team training	Code discussion of support teams receiving: • 8 hours of training on the basic POD design and procedures, referred to as POD 101 • POD IDS User Training, including both general information on using IDS and customized training on role-based functionality
f. Mathematica recruitment staff training	Code discussion of training Mathematica recruitment staff receiving: • Detailed role-based training to prepare them to assist POD callers; the Mathematica survey team provide this training. • Refresher role based training delivered right before the start of recruitment.
g. Social Security processing staff training	Social Security staff training • Detailed role-based training to prepare them to adjust treatment subjects' SSA administrative records under POD rules; SSA will provide this training. • POD Automated System training.
6. Competency-based Certified Work Incentives Counselor certification	Code discussion of the competency-based certification and its three components: • Component 1 – Knowledge Assessment • Component 2 – Case Study Exercise • Component 3 – Benefit Summary and Analysis

EXHIBIT D.3. (CONTINUED)

Contextual Code	Definition and Coding Rules
Technical assistance (TA) and/or unmet TA needs	Code discussion of TA provided by Virginia Commonwealth University, Abt, and SSA; or any unmet TA needs. The Virginia Commonwealth University Site Liaisons will be responsible for monitoring the performance of the POD sites and delivering TA when they identify a need. Code discussion of the TA provided by Abt and Virginia Commonwealth University and site monitoring, including: • Site-specific case reviews (discussing difficult cases as a group) • One-on-one case reviews with individual counselors • File audits of individual participants • TA plans • National video conference calls • Site visits SSA and Abt will also provide policy and operational guidance that will alter/influence how the intervention is implemented. Code discussion of policy or operational guidance provided by SSA or Abt Associates. • Double code with relevant Program Component code.

Note: The coding scheme is subject to change as data collection plans are refined further.

Exhibit D.4. Indicators of implementation context and fidelity of staffing and service delivery in YYYY

Indicator	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	All Sites Combined
Staffing									
Number of work incentives counselors on staff									
Percent work incentives counselors certified at time of hire									
Average number of years since Certified Work Incentives Counselor certification obtained									
Average caseload per full time equivalent work incentives counselor									
Percent of full time equivalent work incentives counselors assigned participants in only one treatment group									
Number of work incentives counselors who have left their position since program began									
Trainings delivered to work incentives counselor staff									
Number of trainings delivered to POD staff									
Percent trainings delivered in-person									
Percent trainings delivered virtually									
Percent trainings self-directed									
Remote service delivery									
Percent of counseling sessions occurring remotely									
Percent of treatment subjects receiving more than half of counseling sessions remotely									

Note: The measures are subject to change as design and data collection plans are refined further.

Exhibit D.5. Indicators of implementation context and fidelity of work incentives counseling in YYYY

Indicator	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	All Sites Combined
Onboarding of new treatment subjects									
Average amount of time to first work incentives counselor contact attempt									
Percent of subjects reached by a work incentives counselor									
Percent of subjects reached by a work incentives counselor who opt out of counseling services									
Develop benefits summary and analyses and work incentives plan									
Percent of clients with benefits planning query before benefits summary and analyses									
Percent of all clients with a benefits summary and analyses									
Percent of employed clients with a benefits summary and analyses									
Percent of clients with an employment goal with a benefits summary and analyses									
Percent of non-working clients without employment goals with a benefits summary and analyses									
Percent of all clients with a work incentives plan									
Percent of employed clients with a work incentives plan									
Percent of clients with an employment goal with a work incentives plan									
Percent of non-working clients without employment goals with a work incentives plan									
Average duration between work incentives plan delivery and next contact									
Deliver ongoing work incentives counseling									
Average number and duration of contacts per work incentives counselor client last quarter									
Average number of e-mail contacts per client									
Average number of phone or in-person contacts per client									
Average duration of contacts per client									
Average number of employment-support referrals last quarter									
Average number of employment-service referrals last quarter									
Average number of referrals to Employment Network									
Average number of referrals to VR									
Average number of referrals to American Job Center									
Percent with benefit adjustment who received counseling within one month of initial benefit adjustment under POD									

Note: The measures are subject to change as design and data collection plans are refined further.

Exhibit D.6. Indicators of implementation context and fidelity of transitioning participants out of POD in YYYY

Indicator	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	All Sites Combined
Percent of participants who transitioned out of POD									
Percent who transition out of POD because participant requested to withdraw									
Percent who transition out of POD because of medical termination									
Percent who transition out of POD because participant is ineligible									
Percent who transition out of POD because of T2 POD earnings termination									
Percent who transition out of POD for some other reason									
Percent of subjects who transitioned out of POD contacted within specified time frame									
Percent of T2s with POD earnings termination contacted within 4 months of scheduled end date									
Percent of withdrawn subjects with transition completed by indicated date									

Note: The measures are subject to change as design and data collection plans are refined further.

Exhibit D.7. Indicators of implementation context and fidelity of benefit adjustment in YYYY

Indicator	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	All Sites Combined
POD Benefit Adjustment									
Percent of subjects known to be in POD offset as of October YYYY									
Percent in POD offset with full benefit offset in October YYYY									
Percent in POD offset receiving less than 50% of full benefit amount in October YYYY									
Percent in POD offset receiving 50-75% of full benefit amount in October YYYY									
Percent in POD offset receiving more than 75% of full benefit amount in October YYYY									
Annual Benefit Reconciliation									
Percent who used the POD offset in YYYY with complete end of year reconciliation documentation submitted timely to SSA									
Percent of YYYY POD offset users who were overpaid in that year									
Percent of YYYY POD offset users who were correctly paid in that year									
Percent of YYYY POD offset users who were underpaid in that year									
Benefit Adjustment Appeals									
Percent of beneficiary-offset months in YYYY for which beneficiaries filed reconsiderations to dispute monthly offset adjustment									
Average time from monthly reconsideration filing to resolution									
Percent of monthly reconsiderations leading to adjustments									
Percent of beneficiaries who used the offset in YYYY who filed reconsiderations to dispute annual adjustment									
Percent of annual reconsiderations leading to adjustments									
Average time from annual reconsideration filing to resolution									

Note: The measures are subject to change as design and data collection plans are refined further.

APPENDIX E

SUPPLEMENT TO CHAPTER VIII

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

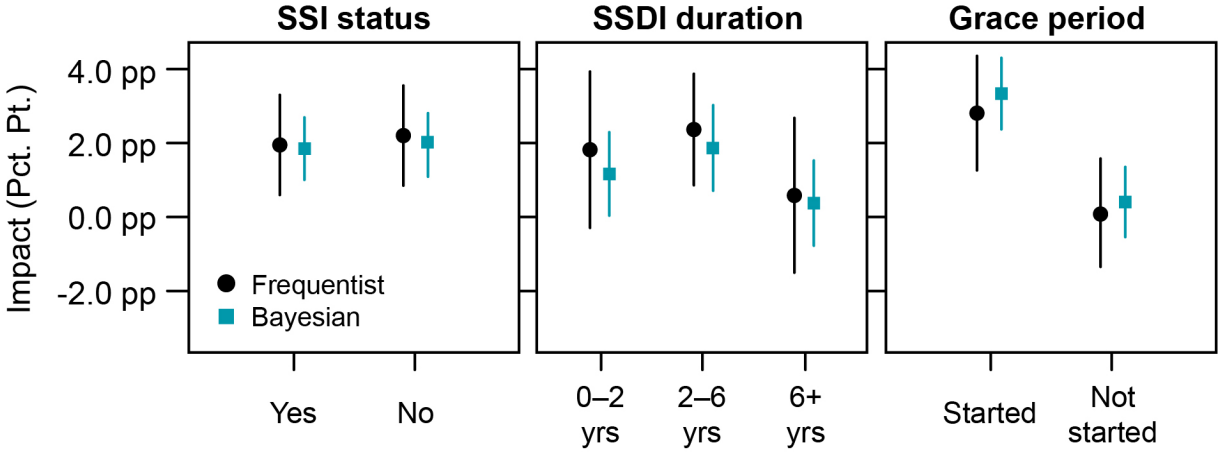
A. Bayesian options for special topic reports

The limitations of traditional, frequentist approaches to subgroup analyses stem from how the standard regression approach is implemented: we can account for some similarities between subgroups through the covariates (X_i), but we otherwise estimate essentially separate impacts for each subgroup. For a special topics report, we plan to explore a Bayesian approach that addresses these issues. The Bayesian approach addresses these limitations by partially pooling information—or borrowing strength—across subgroups. This unified approach will enable us to produce subgroup impact estimates that are more precise and predictive (Gelman et al. 2014); borrowing strength reduces the mean-square error of each subgroup impact estimate.

The potential gains in inference from using a Bayesian framework come at the cost of added assumptions, but we think that these assumptions needed for the POD subgroup analyses are relatively mild. Specifically, the Bayesian framework requires leveraging prior information to achieve better estimates. For the proposed subgroup analyses, we need to simply specify that the impact for younger beneficiaries has *some* correlation with the impact for older beneficiaries, which enables us to borrow information across the groups. We do not specify the exact degree of that correlation. Instead, we estimate it using what we observe for POD subjects, thereby letting the data dictate the extent to which, say, what we find for older beneficiaries influences our impact estimates for younger beneficiaries. We could also refine this approach by (1) specifying that the degree of correlation differs across different subsets of the POD subject pool (to the extent established by the data) and (2) establishing a bound on the likely range of subgroup impacts. The latter assumption would reduce the influence of an outlier result.

In the special topics report, we would present the Bayesian subgroup impact estimates and confidence intervals alongside the main (frequentist) estimates to show how the different approaches change the estimates, precision, and conclusions. Exhibit E.1 illustrates, with fabricated data, how we would present such estimates for subgroups defined by SSI status, SSDI duration, and Grace Period status.

Exhibit E.1. Comparing Bayesian and frequentist subgroup impact estimates for substantive employment (illustrative examples)



Source: SSA program data and baseline survey.

B. Estimating the probability that the intervention is truly effective when the estimated impact is statistically significant

To inform policy, we would like to use the impact findings to state the likelihood that the intervention is truly effective. As discussed in Chapter VIII, researchers sometimes misinterpret the *p*-value as the probability that the true impact is zero, given what we observe in our data. However, we can draw on information from other studies to estimate such a probability. To do this, we have to know the impact and standard error estimates from our study (the same information used to calculate a *p*-value or a confidence interval); the smallest impact the intervention must achieve in order to be considered effective; the proportion of similar interventions that are effective for a given outcome, based on previous research. We will assess the sensitivity of our estimated probabilities to different definitions of *effective* and *similar*.

Ideally we would estimate this probability for each of the primary outcomes, but some of the primary outcomes may not be measured in comparable studies. For example, the BOND, Ticket to Work, and Accelerated Benefits evaluations are relevant for estimating this probability of a true program effect, but they do not all estimate impacts for a measure comparable to our measure of substantive employment (defined as earnings above SGA). Conversely, employment is a secondary outcome, but because it is measured consistently across relevant studies and still important, it would be a candidate to include in this analysis.

To illustrate how these three pieces of information contribute to our assessment of the probability that the benefit offset is effective, we combine them all into an example figure (Exhibit E.2). In this artificial example, we show (in bold black) an impact of the benefit offset on employment of 2.5 percentage points with a 95 percent confidence interval ranging from 1 to 4. The light blue circles in the figure represent impacts estimated in (hypothetical) past studies of similar interventions. By *similar* we mean other interventions that attempted to increase employment for SSDI recipients. The dashed horizontal line represents the threshold for being deemed effective—an impact of 2 percentage points.

In this artificial example, there appears to be a good chance the benefit offset is truly effective. The point estimate is above the yellow line, the lower bound of the confidence interval is above zero, and past research shows that it is not unusual to find impacts on employment that are greater than 2 percentage points (9 of the 20 previous impact estimates are above 2 percentage points).

By way of contrast, Exhibit E.3 shows an example in which there is less chance that the benefit offset is truly effective at increasing employment rates. In this example, the estimated impact and confidence interval are the same as the first example, but the impacts estimated in (hypothetical) past studies show that it is very unusual for programs to have an impact large enough to be deemed effective (only 2 of the 20 previous impact estimates are above 2 percentage points). In this example, we would need a much more precisely estimated impact to be confident that it is a truly effective program—rather than random noise—that resulted in the point estimate being above 2 percentage points.

Exhibit E.2. Assessing the probability POD is truly effective, Example 1

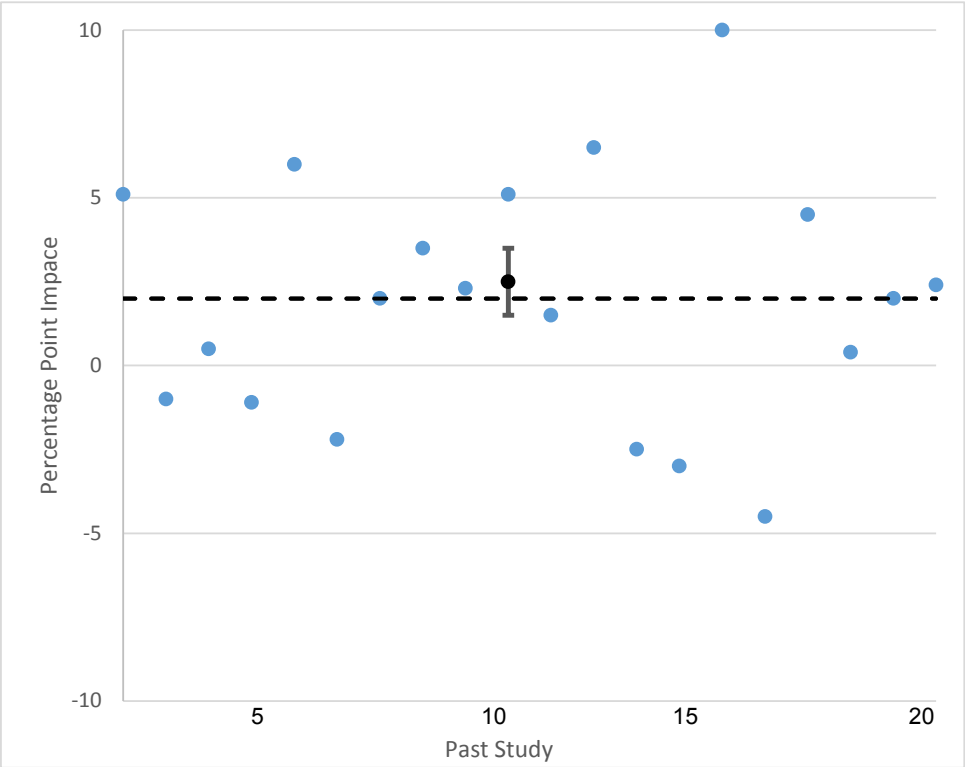
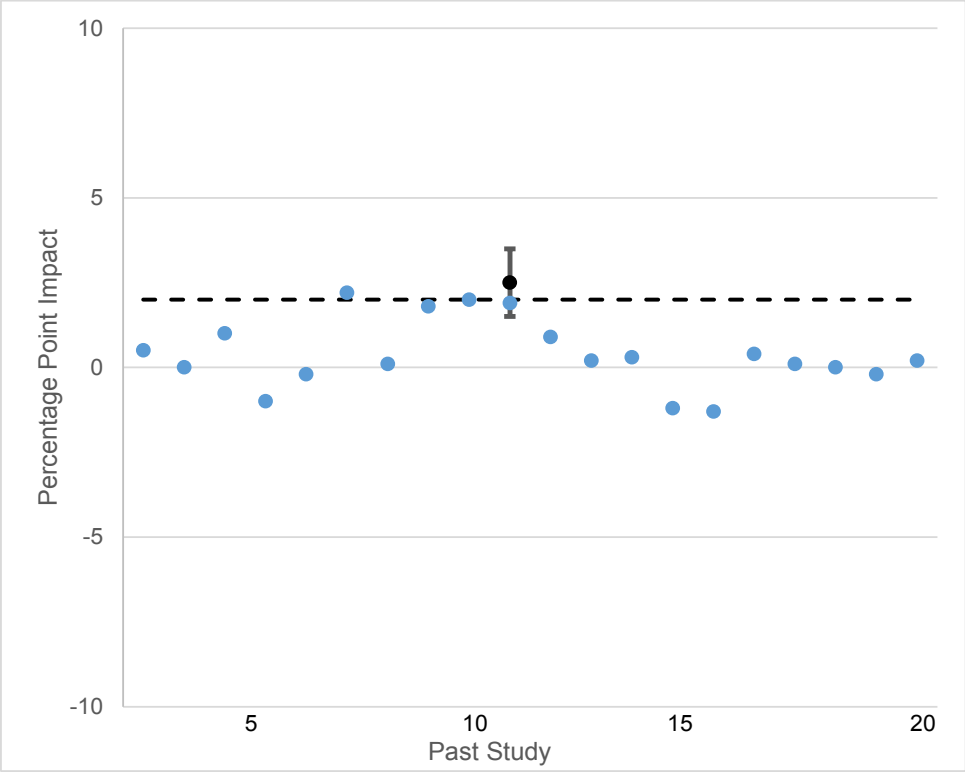


Exhibit E.3. Assessing the probability POD is truly effective, Example 2



We can use the information displayed in Exhibits E.2 and E.3 to calculate the probability that we would like to know: that the benefit offset truly increases employment when the estimated impact is statistically significant. The probability that the benefit offset truly increases employment when the estimated impact is statistically significant is 1 minus the probability of a false discovery. The false discovery rate (FDR) is the fraction of all statistically significant impact *estimates* in which the *true* impact is zero (Benjamini and Hochberg 1995; Storey 2003; Colquhoun 2014). This fraction is stated in Equation (1), where R is the number of rejected null hypotheses and V is the number of falsely rejected null hypotheses.[1] For example, if the null hypothesis is that the true impact of POD is zero, then the null is falsely rejected when (1) the true impact really is zero and (2) the estimated impact is statistically significant. The null is correctly rejected when the null is actually not true (that is, when the true impact is not zero).

(1) $$FDR = E\left[\frac{V}{R}\middle| R>0\right]$$

The FDR can also be expressed as in Equation (2),[2] where the symbol H_0 represent the event that the null hypothesis is true (for example, the true impact of POD on employment is zero), the symbol H_1 represents the event that a specific alternative hypothesis is true (for example, the true impact of POD on employment is 2 percentage points), *reject* means that the null hypothesis is rejected (for example, because the impact estimate is statistically significant), the symbol α is the significance level used in hypothesis testing (for example, 5 percent), and *power* is the statistical power to detect a specific impact.

(2) $$P\left(H_0\middle|reject\right) = \frac{P\left(H_0\right)*P\left(reject\middle|H_0\right)}{P\left(reject\middle|H_0\right)*P\left(H_0\right)+P\left(reject\middle|H_1\right)*P\left(H_1\right)}$$

$$\frac{P\left(H_0\right)*\alpha}{\alpha*P\left(H_0\right)+power*P\left(H_1\right)}$$

We can also calculate the probability that an intervention is truly effective when the estimated impact is statistically significant. Equation (3) provides the formula for this probability.

(3) $$P\left(H_1\middle|reject\right) = \frac{P\left(H_1\right)*power}{power*P\left(H_1\right)+\alpha*P\left(H_0\right)}$$

The quantity $P\left(H_1\right)$ can be estimated using data. In our example figures in Chapter VII, $P\left(H_1\right)$ is estimated to be the proportion of black circles above the gold line (0.45 for Example 1; 0.10 for Example 2). For these two examples, we assume that *power* is 80 percent and that α

[1] We use the definition of the FDR proposed by Storey (2003) in which the FDR is defined only when R > 0.

[2] Storey (2003) and Colquhoun (2014) present formulas similar to Equations (2) and (3).

is 5 percent. Substituting these values into Equation (3) yields Equation (4) for Example 1 and Equation (5) for Example 2.

$$(4)\ P(H_1 \mid reject) = \frac{0.45 * 0.8}{0.8 * 0.45 + 0.05 * 0.55} = 0.93$$

$$(5)\ P(H_1 \mid reject) = \frac{0.10 * 0.8}{0.8 * 0.10 + 0.05 * 0.90} = 0.64$$

For Example 1 (Exhibit E.2), the probability that the benefit offset is truly effective given that the impact is statistically significant is 93 percent. For Example 2, that probability is 64 percent. These probabilities illustrate the point made by the American Statistical Association statement on *p*-values—the *p*-value in and of itself does not tell us the probability that an impact is due to chance. The impact, standard error, and *p*-values are the same in these two examples, yet the probability that the impact is real (that is, greater than 2 percentage points as opposed to being the result of random chance) differs substantially between the two examples.

When interpreting findings, we will include a table showing estimates of the probability that each statistically significant impact is due to a true effect of the benefit offset, as opposed to random chance (Exhibit E.4). Because these estimates depend on subjective judgment regarding which past studies are relevant to include when calculating the proportion of past studies in which the intervention was effective, we will show multiple estimates of this probability for each statistically significant impact to assess sensitivity to subjective judgment.

Exhibit E.4. Assessing the probability that significant impacts are truly greater than or equal to the MDI

Outcome	Contrast	MDI	Assumed prevalence of impacts greater than MDI	Probability that the true impact of POD is greater than MDI
Employment rate	T1 versus C	2.2	0.45 0.10	0.93 0.64

C = control; MDI = minimum detectable impact; T = treatment.