BOND Implementation and Evaluation

Final Evaluation Report
Volume 1

Deliverable24e.2

October 18, 2018

Prepared by:

Daniel Gubits
David Stapleton
Stephen Bell
Michelle Wood
Denise Hoffman
Sarah Croake
David R. Mann

Judy Geyer
David Greenberg
Austin Nichols
Andrew McGuirk
Meg Carroll
Utsav Kattel
David Judkins

Abt
ASSOCIATES

MATHEMATICA
Policy Research, Inc.

Table of Contents

List of Exhibits

Acronyms Used in This Report

AEE	Annual Earnings Estimate	**FTE**	Full-Time Equivalent
AIME	Average Indexed Monthly Earnings	**GP**	Grace Period
BODS	BOND Operations Data System	**I&R**	Information and Referral
BOND	Benefit Offset National Demonstration	**IRP**	Initial Reinstatement Period
BPP	BOND Participation Period	**IRS**	Internal Revenue Service
BS&A	Benefits Summary and Analysis	**IRWE**	Impairment-Related Work Expenses
BSAS	BOND Stand Alone System	**MBI**	Medicaid Buy-In
BTS	Beneficiary Tracking System	**MBR**	Master Beneficiary Record
BYA	BOND Yearly Amount	**MEF**	Master Earnings File
CDR	Continuing Disability Review	**METB**	Marginal Excess Tax Burden
CPE	Centralized Post Entitlement	**ORDES**	Office of Research, Demonstration and Employment Support
CWIC	Community Work Incentive Coordinators	**PHUS**	Payment History Update System
DAC	Disabled Adult Child	**SGA**	Substantial Gainful Activity
DAF	Disability Analysis File	**SSA**	Social Security Administration
DBAD	Master Beneficiary Record, Disabled Beneficiary and Dependent files	**SSDI**	Social Security Disability Insurance
DWB	Disabled Widow/Widowers Benefits	**SSI**	Supplemental Security Income
EN	Employment Network	**SSR**	Supplemental Security Record
EOYR	End of Year Reconciliation	**TWP**	Trial Work Period
EPE	Extended Period of Eligibility	**VR**	Vocational Rehabilitation
EXR	Expedited Reinstatement	**WIC**	Work Incentive Counseling, Or Counselor
EWIC	Enhanced Work Incentives Counseling, Or Counselor	**WIPA**	Work Incentives Planning and Assistance

Terminology

1. **BOND subjects:** Beneficiaries assigned to any of the five BOND treatment or control groups, at either stage (see **Exhibit 1-1**). Terms for subjects in specific groups are as follows:

   a. **Treatment subjects:** All subjects offered the use of the benefit offset, including:

      i. **T1 subjects** or **Stage 1 treatment subjects:** Those offered the offset at Stage 1.

      ii. **Stage 2 treatment subjects:** Those offered the offset at Stage 2, including:

         (1) **T21 subjects:** Stage 2 volunteers offered the offset, but not offered enhanced work-incentives counseling.

         (2) **T22 subjects:** Stage 2 volunteers offered both the offset and enhanced work-incentives counseling.

   b. **Control subjects:** Those whose benefits will continue to be determined by current law.

      i. **C1 subjects:** Those assigned to the Stage 1 control group.

      ii. **C2 subjects:** Stage 2 volunteers assigned to the Stage 2 control group.

2. **BOND users:** Those treatment subjects who take up a BOND treatment. These include:

   a. **Offset users** – All treatment subjects who have their benefits reduced by the offset.

   b. **EWIC users** – All treatment subjects who use EWIC services. They can only be subjects in the T22 group.

   c. **WIC users** – All treatment subjects who use WIC services. They can be subjects in the T1 or T21 groups.

Executive Summary

As part of the Ticket to Work (TTW) and Work Incentives Improvement Act of 1999, Congress directed the Social Security Administration (SSA) to test alternative Social Security Disability Insurance (SSDI) work rules designed to increase the incentive for SSDI beneficiaries to work and reduce the total amount of SSDI benefits paid to beneficiaries. In response, SSA has undertaken the Benefit Offset National Demonstration (BOND), a random assignment test of alternative SSDI program rules governing work and other supports. BOND tests a $1 for $2 benefit offset applied to annual earnings above the BOND Yearly Amount (BYA)—the annual equivalent of SSDI's substantial gainful activity amount. As a result, beneficiaries in the treatment group are able to retain some of their monthly cash benefits while earning more than BYA. The benefit offset reduces yearly SSDI benefits by $1 in SSDI benefits for every $2 in annual earnings above BYA (in other words, reduces yearly SSDI benefits by half the amount that annual earnings exceed the BYA threshold).

BOND includes two stages. The purpose of Stage 1 is to learn how a national benefit offset would affect earnings and program outcomes for the entire SSDI population. To achieve this goal, Stage 1 uses two-way random assignment into an offset treatment group (with standard work incentives counseling) or to a current-law control group.

The purposes of Stage 2 are to (i) learn more about impacts on those beneficiaries thought most likely to use the offset (recruited and informed volunteers who are not also receiving Supplemental Security Income) and (ii) to determine the extent to which enhancements to counseling services (enhanced work incentives counseling, or EWIC) affect impacts compared to standard work incentives counseling (WIC). To achieve these goals, Stage 2 uses three-way random assignment into an offset-plus-WIC group, an offset-plus-EWIC group, and a current law control group.

This Final Evaluation Report presents findings about the implementation of BOND through the sixth year of implementation (2016), estimates of impacts on earnings and benefit outcomes from the first five years of implementation (through 2015), and the results of benefit-cost analyses.

Summary of Key Findings

Impacts on Earnings and SSDI Benefits

The analysis finds no confirmatory[1] evidence of an impact of the benefit offset on average earnings in either the nationally-representative Stage 1 or in the Stage 2 sample of volunteers.

In contrast, the analysis finds confirmatory evidence that, relative to current law, the benefit offset policy increased the average amount of SSDI benefits due to beneficiaries over five years. In the nationally-representative Stage 1, the positive impact on SSDI benefits was $143 per year (or about $12 per

---

[1] The evaluation seeks conclusive evidence of impacts of the benefit offset policy on beneficiaries' earnings and SSDI benefits over the five-year follow-up period. It examines other impacts using a less demanding standard of evidence for concluding that the offset policy had an effect. If found statistically significant, a positive effect on average earnings or a negative effect on SSDI benefits would confirm the intervention's success and hence constitute "confirmatory" evidence.

month)—an increase of slightly more than 1 percent of the current-law average benefits. In Stage 2's sample of study volunteers, the positive impacts on SSDI benefits were larger than in Stage 1, about 4 percent of current-law average benefits.

Underlying these positive impacts on SSDI benefits are factors that changed benefits in opposite directions. On the one hand, two factors pushed average SSDI benefits upward. First, under the offset policy, SSA pays partial benefits to those who would have had zero benefits due under current law (in effect, a windfall to those who would already engage in substantial gainful activity under current law). Second, economic theory predicts that those beneficiaries receiving windfall income will respond by reducing their earnings somewhat, thereby further increasing average SSDI benefits. The evaluation finds evidence in Stage 1 of this reduction in earnings by at least some of those who would engage in substantial gainful activity under current law.

On the other hand, economic theory also predicts that the benefit offset will increase the proportion of beneficiaries earning more than BYA. Those who are induced to increase their earnings above BYA receive partial SSDI benefits under the offset policy rather than full benefits. This serves to decrease average SSDI benefits. Consistent with this theory, the evaluation finds evidence in both stages that the benefit offset increased the share with earnings above BYA (relative-to-control-mean, by 7 percent in Stage 1 and 23 percent in Stage 2).

The evidence of positive impacts on SSDI benefits shows that the two factors increasing benefits outweighed the one factor serving to decrease benefits. In fact, the magnitude of the increase in the percentage earning above BYA was far from large enough in either stage to reduce average SSDI benefits for all treatment beneficiaries.[2]

In addition to examining effects of the benefit offset, Stage 2 tested the extent to which enhancements to counseling combined with the offset produced improved results compared to the offset with standard counseling. The analysis finds virtually no evidence of impacts from counseling enhancements on any earnings-related or benefit-related outcome.

Benefit-Cost Analysis Results

For the nationally representative Stage 1 sample, the benefit-cost analysis found a net social cost of the BOND offset. The very small estimated increases in earnings were not sufficient to offset the deadweight loss from increases in taxes needed to fund larger SSDI benefit payments. Distributional effects were much larger, with SSDI beneficiaries gaining income by receiving larger SSDI benefits and countervailing losses to the Disability Insurance Trust Fund. The benefit-cost analysis for Stage 2 shows that the offset policy combined with standard work incentives counseling has a net social benefit in the small subpopulation of beneficiaries who volunteer for the demonstration. In contrast, the offset policy combined with enhanced work counseling has a net social loss, a result largely due to higher counseling costs.

---

[2]   By a rough estimate, the Stage 1 increase in the percentage earning above BYA would need to be 30 times as large as the observed impact in 2014 to drive the impact on SSDI benefits to zero. The corresponding multiple for each Stage 2 treatment group is close to 10.

Offset Use

A small minority of beneficiaries in each treatment group used the offset, although use increased each year in both demonstration stages. In Stage 1, 3.6 percent of treatment subjects used the offset in any year during the first five years and 2.2 percent used the offset in the final year examined, 2015. Stage 2 offset use was higher, as expected (about 15 percent of treatment subjects used the offset in any year during the first five years and 10 percent used the offset in 2015). The evaluation finds no evidence in Stage 2 that EWIC resulted in higher benefit offset use than WIC.

Potential Explanations for Why Offset Use and Earnings Effects Were Not Larger

The evaluation findings offer four potential explanations for why offset use and the effect on percentage earning above BYA were not larger in magnitude. One possible explanation is the limited ***work capacity*** of most beneficiaries. Because of the nature of the SSDI eligibility criteria, only a fairly small minority of beneficiaries were expected to use the offset.

A second possible explanation is that the increase in the ***incentive*** to earn more than BYA, although strong, was not strong enough to induce some beneficiaries with the capacity to earn more than BYA to actually do so. The offset imposes an implicit 50 percent tax on earnings. For some, this implicit tax, perhaps along with potential tax increases and reductions in other benefits, may make the increase in net income from increased earnings smaller than the opportunity cost of giving up other activities.

A third possible explanation stems from the ***complexity*** of both the offset and the current law rules, which muddles treatment subject understanding of the difference in the offset and current-law incentives.

The final possible explanation interacts with the complexity of the rules: that the conditions established by the BOND ***implementation*** may have led to smaller impacts than would be the case under a permanent national benefit offset policy. The findings for Stage 1 suggest that outreach and information to treatment subjects was not sufficient for them to achieve the level of understanding of the offset rules that control subjects have of current law. Further, long delays in making first benefit adjustments may have depressed understanding of the offset rules.

It is possible that the impact on the proportion earning above BYA might have been somewhat larger in the nationally-representative Stage 1 had outreach to treatment subjects been more robust and benefit adjustments been made in a more timely manner. However, there is nothing in the evidence to suggest that the magnitude of the effect under an established national program would reach the effect found for volunteers in Stage 2 (which at 23 percent is more than three times greater than the 7 percent relative effect found in Stage 1). Further, the results from Stage 2 imply that even a 23 percent effect would fall far short of what is required to reduce total SSDI benefits—even before consideration of induced entry.

Therefore, the evidence from BOND shows that a national policy that reduces SSDI benefits by $1 for every $2 in earnings above the substantial gainful activity threshold would not reduce the total amount of SSDI benefits owed to beneficiaries.

1.   Introduction

Administered by the U.S. Social Security Administration (SSA), Social Security Disability Insurance (SSDI) is the nation's primary earnings-replacement program for workers who become unable to work substantially due to long-term or terminal physical or mental conditions. In 2015, the SSDI program paid $143 billion in cash benefits to approximately 9 million disabled workers and 2 million of their spouses and disabled children (Burke 2016; SSA 2017a).

Several factors—low benefit termination rates, lower ages at disability onset, and a growing number of claims—have contributed to substantial growth in the SSDI caseload and in program costs since 1991 (Morton 2013; Autor and Duggan 2006).[3] The Social Security Board of Trustees project that the DI Trust Fund[4] will be exhausted by 2032 (SSA 2018). Although the SSDI program offers work incentives to encourage beneficiaries to attempt to return to work, under current law beneficiaries lose their entire disability benefit after a sustained period of substantial earnings ($1,090 per month in 2015). It is widely presumed that this "cash cliff" discourages work and decreases exit from the program. More work among SSDI beneficiaries would lead to reduced benefits and higher payroll tax payments, which would be an aid in addressing the problem of sustaining the DI Trust Fund.

SSA has a broad initiative to identify and test policies that have the potential to help SSDI beneficiaries increase earnings and reduce reliance on SSDI benefits, thereby reducing the program's total cost.[5] A leading part of that initiative is the Benefit Offset National Demonstration (BOND). As part of the Ticket to Work and Work Incentives Improvement Act of 1999,[6] Congress asked SSA to analyze the costs and benefits of replacing the SSDI cash cliff with a more gradual reduction in benefits, a "ramp" when earnings exceed substantial gainful activity (SGA). BOND tests this alternative policy. SSA, in conjunction with several contractors led by Abt Associates, developed the infrastructure and supports required to implement the $1 for $2 benefit offset analyzed in BOND.

---

[3]   The financial outlook of the DI Trust Fund has improved somewhat in recent years. In their most recent annual report, the Social Security Board of Trustees note that disability applications have been declining steadily since 2010 and that the total number of disabled-worker beneficiaries in current payment status has been falling since 2014 (SSA 2018). These trends have led to a change in the projected date of trust fund depletion, from 2023 (projected in 2016) to the current projection of 2032 (SSA 2016a; SSA 2018).

[4]   Social Security taxes and other income are deposited in the Disability Insurance (DI) Trust Fund and SSDI benefits are paid from it. The purposes for which the DI Trust Fund can be used are to pay benefits and program administrative costs. The Annual Report of the Social Security Board of Trustees analyzes the actuarial status of the Old Age and Survivor's Insurance (OASI) and Disability Insurance (DI) trust funds (SSA 2017b)

[5]   See SSA's website "What are demonstration projects?" (SSA n.d., accessed August 30, 2017) for more information on SSA's other demonstration projects that test SSDI program changes. Section 234 of the Social Security Act, as amended in the Bipartisan Budget Act of 2015 gives SSA authority to conduct research and demonstration projects through December 31, 2022.

[6]   Ticket to Work and Work Incentives Improvement Act of 1999, Public Law 106-170, Section 302, 106th Congress, codified at U.S. Code 42 (1999) §434.

This is the final report for the BOND evaluation. The balance of this chapter proceeds as follows. Section 1.1 describes the BOND policy tests. Section 1.2 describes the evaluation. Section 1.3 describes the sites and the sample. Finally Section 1.4 describes the organization of the balance of this document.

The BOND Evaluation Team

Abt Associates, in partnership with 25 other organizations, implemented and has now concluded the bulk of the evaluation for the Benefit Offset National Demonstration (BOND) under contract to the U.S. Social Security Administration. To ensure the objectivity of the evaluation, separate teams conduct the implementation (the "Implementation Team") and evaluation (the "Evaluation Team") components of the project. The current report reflects exclusively the views of the evaluation team, led by Evaluation Co-Directors Stephen Bell of Abt Associates and David Stapleton of Mathematica Policy Research. Beginning in April 2018 Daniel Gubits has served as Co-Director of Evaluation at Abt Associates. These individuals have no role in implementing or overseeing the BOND intervention they are studying, nor do any members of their evaluation team. The Abt Project Director (Michelle Wood) is responsible for coordinating the implementation and evaluation efforts, including, respectively, managing the day-to-day operations of the project and overseeing the effective and efficient implementation of the BOND design. Within this structure, full authority over and responsibility for the content of all evaluation reports rests with the Evaluation Co-Directors.

1.1.   The BOND Policy Test

Under current-law program rules, SSDI beneficiaries lose all SSDI benefits after a sustained period of substantial earnings and risk potential loss of other (non-SSDI) benefits.[7] Specifically, SSDI benefits are lost if, after completing a nine-month Trial Work Period (TWP) and a three-month grace period (GP), a beneficiary's countable monthly earnings exceed the monthly Substantial Gainful Activity (SGA) amount.[8] For the first 36 months after the TWP is completed (except in the three GP months), SSA suspends benefits when earnings exceed the SGA amount. After these 36 months, SSA terminates SSDI entitlement if any month's earnings exceed the SGA amount. Economic theory suggests that the cash cliff discourages some beneficiaries from working at all and encourages those who work and could earn above the SGA amount to keep their earnings below that level.

In contrast, for beneficiaries subject to the alternative rules tested in BOND, SSA replaces the cash cliff with a ramp (i.e., the benefit offset). Under the new rules, benefits are partially reduced as earnings increase. The benefit offset reduces benefits by $1 for every additional $2 in earnings above an SGA-equivalent threshold, based on annual earnings—the BOND yearly amount (BYA). The policy objective

---

[7]   Other benefits include Medicare for those on the rolls for at least 24 months. These benefits are extended for a lengthy period following suspension of SSDI benefits, but not indefinitely. Some SSDI beneficiaries also receive Supplemental Security Income (SSI), Medicaid, or other public or private benefits that can be reduced or eliminated as earnings increase.

[8]   In 2015, the final year for which this report analyzes the impacts of the benefit offset, the SGA amount was $1,090 per month for non-blind beneficiaries and $1,820 per month for blind beneficiaries.

of the offset is to encourage more beneficiaries who can earn above the SGA level to increase their earnings and reduce their reliance on benefits. The benefit offset is expected to increase the earnings of some who would otherwise earn less than the SGA amount or might not work at all. If such individuals engage in SGA under the benefits offset rules, their SSDI benefits will be partially reduced. The reduction from full benefits to partial benefits for these beneficiaries creates the possibility that benefit offset policy could reduce the total cost of the SSDI program.

There are, however, opposing effects on earnings and benefits for a small, but important, subgroup of beneficiaries: those who would earn more than the SGA level under current law. Many of those beneficiaries who would regularly earn more than the SGA level under current law (and so have suspended benefits) will be eligible for a partial SSDI benefit under the offset rules, in effect providing them with a windfall for no change in behavior. In addition, some beneficiaries who would earn more than SGA under current law will be expected to choose to earn less when subject to the benefit offset. Such individuals can achieve the same level of income—earnings plus DI benefits—with lower earnings.

Thus, the net impact on mean earnings and benefits of all beneficiaries depends on the size of the impacts for beneficiaries who would not earn more than the SGA level under current law relative to the size of the impacts for those who would earn more than SGA (summarized in Exhibit 1-1).[9]

Exhibit 1-1.   Hypothesized effects of the $1 for $2 Benefit Offset on Earnings and SSDI Benefits

Beneficiaries Who…	Hypothesized Effect on Earnings		Hypothesized Effects on SSDI Benefit	
Would earn below the SGA level under current law	↑ Increase in average earnings	Some of these beneficiaries would increase their earnings in response to the removal of the "cash cliff".	↓ Decrease in average SSDI benefits	Some of these beneficiaries would choose to earn above BYA[a], which would reduce their benefits from full to partial.
Would earn above the SGA level under current law	↓ Decrease in average earnings	Some of these beneficiaries would reduce earnings in response to income and substitution effects created by the offset "ramp".	↑ Increase in average SSDI benefits	Most of these beneficiaries would go from zero benefits to partial benefits.
All beneficiaries	? Ambiguous	Depends on relative strength of first two effects.	? Ambiguous	Depends on relative strength of first two effects.

Note: [a] BYA is the acronym for "BOND Yearly Amount", the annual SGA-equivalent threshold used by the BOND benefit offset rules. BYA equals 12 times the monthly SGA threshold.

---

[9]   BOND tests the offer of the benefit offset only to beneficiaries already on the SSDI rolls. Under a national program, the benefit offset might induce some workers who are not on the rolls to enter SSDI. Maestes et al. (2010) reviews the extensive literature and evidence that increases in program generosity induce entry and reduce the earnings of program entrants. After considerable consultation with experts, SSA determined that it would not be feasible for BOND to measure induced entry using an experimental research design (Tuma 2001). Nonetheless, induced entry remains an important issue and SSA plans to assess the likely extent of induced entry through separate non-experimental research.

For a BOND treatment subject, full understanding of the offset involves understanding how his or her combined income from SSDI benefits and earnings will vary with different levels of earnings. This is a complex relationship, relatively difficult to absorb from either a graphical or narrative presentation. Lessons from the Benefit Offset Pilot Demonstration[10] suggested that treatment subjects were likely to have difficulty achieving this full understanding. SSA believed that additional work incentives counseling might facilitate use of the offset. To address this concern, SSA incorporated the enhanced work incentives counseling (EWIC) component into the Stage 2 test of the offset. Stapleton et al. (2010) notes that EWIC is expected to increase the impact of the offset by improving beneficiary understanding of how higher earnings will affect SSDI and other benefits. EWIC was also intended to help beneficiaries access the medical treatments, employment supports, and job search assistance they might need to address other obstacles. In addition, by establishing a helpful and trusting relationship with beneficiaries, EWIC was expected to overcome skepticism and distrust, making it more likely that treatment subjects would take advantage of the benefit offset to increase their earnings and total income.

1.2.   The BOND Evaluation

Abt Associates and Mathematica Policy Research are conducting a comprehensive evaluation of BOND. The evaluation used random assignment to test the $1 for $2 benefit offset. BOND includes two stages.

- *Stage 1* examines how a national benefit offset would affect earnings and program outcomes for the entire SSDI population. In this stage, the demonstration randomly assigned beneficiaries into either a treatment group "T1" (subject to benefit offset rules and offered Work Incentives Counseling [WIC]) or a current-law control group "C1".

- *Stage 2* tests the impact of the offset for those expected to be most likely to use the offset—recruited and informed volunteers. Stage 2 also analyzes the extent to which enhanced counseling (EWIC) affects impacts. In this stage, the demonstration randomly assigned volunteers into one of three assignment groups: a treatment group "T21" (benefit offset rules and offered WIC), a second treatment group "T22" (benefit offset rules and offered EWIC), or a current-law control group "C2".

Exhibit 1-2 illustrates the relation between Stage 1 and Stage 2 and Exhibit 1-3 gives summary information about the stages and the sample sizes.

---

[10]   Weathers and Hemmeter (2011) summarize findings from the Benefit Offset Pilot Demonstration. The BOPD pilots operated in four states—Connecticut, Utah, Vermont, and Wisconsin from 2005 to 2008. The BOPD provided lessons about implementing a $1 for $2 benefit offset but were not designed to produce estimates of the effects of a national benefit offset program.

Exhibit 1-2.　　Overview of BOND Random Assignment Process

BOND Sample Enrollment

All Eligible DI-Only & Concurrent Beneficiaries in Sites

Stage 1 RA

T1

$1 for $2 Offset
77,101

Eligible DI-Only Beneficiaries: Stage 2 Solicitation Pool
238,070

C1

Control Group
891,429

Recruitment & Consent

RIC

Stage 2 Volunteers

Stage 2 RA

$1 for $2 Offset
4,854

T21

T22

C2

Control Group
4,849

$1 for $2 Offset with Enhanced Work Incentives Counseling
3,041
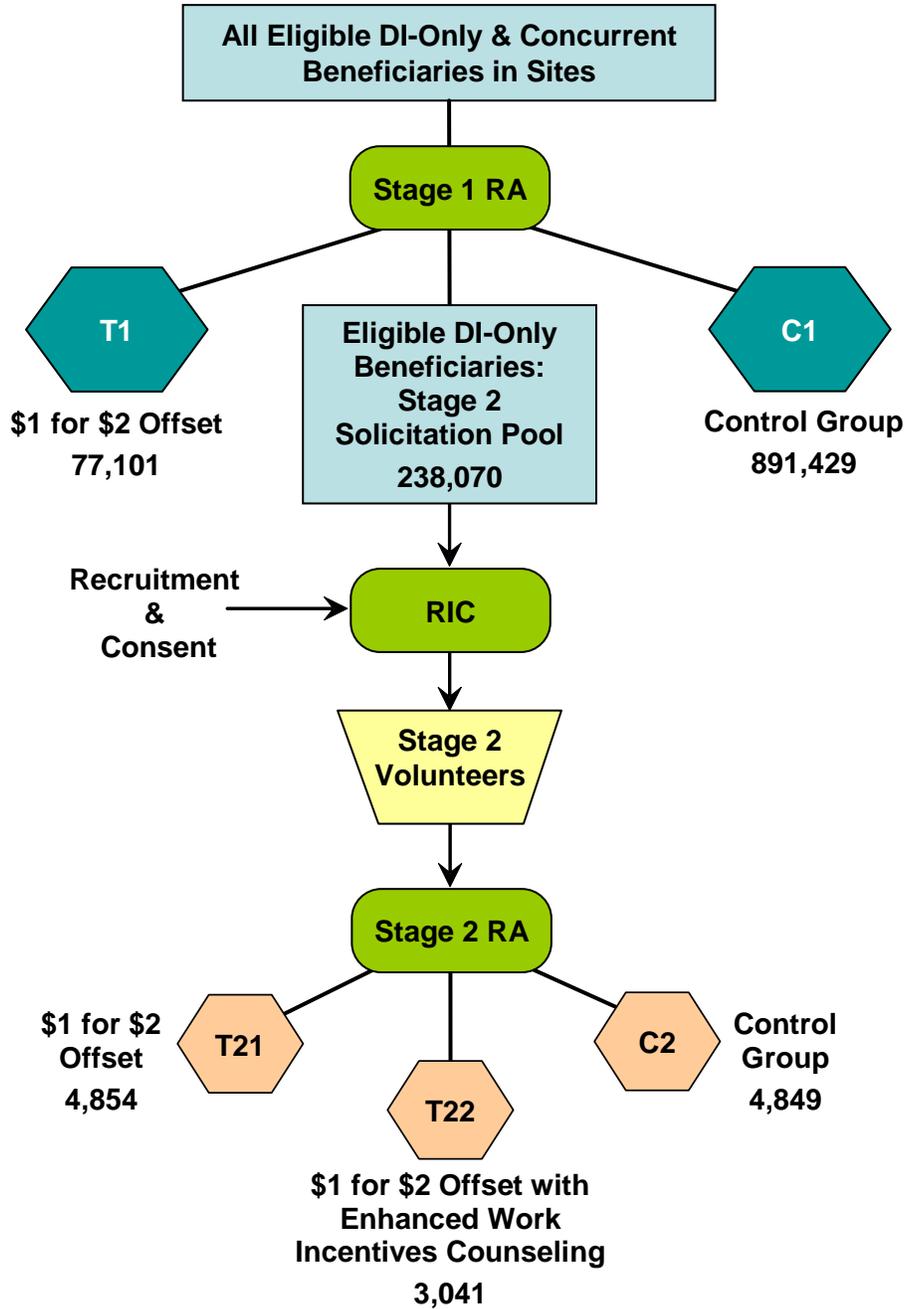
DI = disability insurance; RA = random assignment; RIC = recruitment and informed consent.

Exhibit 1-3.   Attributes of Stage 1 and Stage 2 Samples

	Stage 1	Stage 2
Treatment conditions	T1: Offset, regular WIC	T21: Offset, regular WIC (identical to T1) T22: Offset, Enhanced WIC (EWIC)
Concurrent SSDI/SSI beneficiaries	Included	Excluded at enrollment; later entry to SSI permitted.
Age criterion	Ages 20 to 59 as of May, 2011	Ages 20 to 59 as of May, 2011
Random assignment	Late April 2011	March 1, 2011 to September 28, 2012, with 40 percent of volunteers enrolling in the study in 2011 and 60 percent of volunteers enrolling in 2012
Offset rules first applied	May 2011	Month after random assignment
Weighted percentage of short-duration (36 months or less) beneficiaries	30%	42%
Sample size	C1: 891,429 T1: 77,101	C2: 4,849 T21: 4,854 T22: 3,041

Source: BOND Operations Data System (BODS).

Notes: WIC: Work Incentives Counseling

The total Stage 1 sample size (T1 + C1) is 968,530. The Stage 1 analysis sample excludes (1) subjects who were initially assigned to the sample but were later determined to have died prior to assignment, and (2) any pair of beneficiaries on a common primary record who were assigned to different BOND groups. Examples of the latter include a primary and a disabled adult child (DAC) or two DACs with the same primary beneficiary. See Stapleton et al. (2013) for details on this adjustment. Weights are used to ensure that the BOND subjects who meet the analysis criteria in both the T1 and C1 analysis samples are representative of the national beneficiary population in the month of random assignment. The weighted population size is 6,453,341.

The total Stage 2 sample size (T21 + T22 + C2) is 12,744. The Stage 2 analysis sample excludes 210 beneficiaries who are related to other BOND subjects (e.g., a primary and a DAC or two DACs with the same primary) to avoid contamination effects that might arise from the fact that almost all such beneficiaries (204 of the 210) were assigned to different BOND groups (see Appendix A of Gubits et al. (2017) for details on this adjustment). Because only six of these beneficiaries would have been able to be retained, it was not feasible to replicate the approach used for the Stage 1 analysis (where we were able to include pairs in which both members were assigned to the same group and revise the weights so that impact estimates reflect impacts for all beneficiary pairs with at least one member in Stage 1 (Stapleton et al. 2013).

Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national population of SSDI-only beneficiaries who would volunteer for study enrollment. The weighted Stage 2 sample size is 276,342 (the estimated number of Stage 2-eligible beneficiaries in the nation who would have volunteered had all Stage 2-eligible beneficiaries been offered the opportunity to enroll in the study).

The BOND Evaluation encompasses four studies.

1. The *process analysis* describes the BOND interventions, evaluates demonstration implementation, and assesses the fidelity of the implementation to the original design.

2. The *participation analysis* describes BOND subjects' use of BOND benefits counseling services, work effort, and use of the $1 for $2 benefit offset. The participation analysis examines patterns of participation by subgroups and characteristics that distinguish offset users from other subjects. For Stage 2, the participation analysis also analyzes recruitment, to identify beneficiary characteristics that are associated with volunteering for the demonstration.

3. The *impact analysis* measures the net impact of the opportunity to use the $1 for $2 benefit offset on BOND subjects' employment, earnings, SSDI benefits, and other outcomes. For Stage 2, the impact analysis also measures the impact of EWIC compared to regular work incentives counseling.

4. The ***benefit-cost analysis*** compares the benefits and costs of the $1 for $2 benefit offset policy to current SSDI rules. For Stage 2, the benefit-cost analysis also documents the costs and benefits of EWIC. The analysis examines costs and benefits from multiple perspectives; society as a whole, SSDI beneficiaries, the DI Trust Fund, and other government. The findings from the benefit-cost analysis inform policymakers about the desirability of incorporating the benefit offset into the SSDI program as a national policy.

Two design features of the benefit offset shaped the implementation. First, the benefit offset replaces the monthly SGA calculation with an annualized measure of SGA, referred to as the BOND Yearly Amount (BYA). BYA is equal to 12 times the monthly SGA amount (in 2015, $13,080 for non-blind and $21,840 for blind treatment subjects). The benefit offset reduces benefits by $1 for every $2 in countable annual earnings in excess of the BYA following the completion of the TWP and GP. The annualized measure of SGA can be helpful to beneficiaries who have variable monthly earnings in instances in which annual earnings stay below BYA even though certain months have earnings above SGA. Use of the annualized measure requires changes in SSA's administrative procedures for adjusting benefits. For treatment subjects who have completed the TWP and GP, SSA pays benefits monthly under offset rules, but the monthly payment amount is based on expected annual earnings. In the following calendar year, SSA reconciles payments to those that should have been due given actual countable earnings in the year, based on information provided by the Internal Revenue Service (IRS) and documentation provided by the beneficiary.

Second, BOND treatment subjects can use the benefit offset during a 60-month participation period. For beneficiaries who completed the TWP before random assignment, the BOND participation period starts the month after random assignment. For other beneficiaries, the BOND participation period begins the month after the TWP is completed. In both cases, the offset applies once all GP months are used up. Beneficiaries who do not complete the TWP by September 30, 2017 will lose their opportunity to use the benefit offset.[11] SSA does not terminate SSDI entitlement because of work during the BOND participation period, even if benefits fall to zero because of earnings that are well above the BYA.[12]

1.3.   BOND Sites and Sample

To test the benefit offset policy, Abt Associates randomly selected ten of SSA's 53 Area Offices as study sites. Random selection of sites ensures that the evaluation's findings are nationally representative. Collectively, the demonstration includes 1.2 million of the 9.8 million total SSDI disabled beneficiaries.

1.3.1.   Stage 1 Sample

The Stage 1 sample is a nationally representative cross-section of the SSDI population under age 60 as of May 2011. At each site, all current SSDI beneficiaries between ages 20 and 59 receiving benefits based on disability[13] and who were not part of another SSA demonstration were included in the BOND sample.

---

[11]   Thus, some treatment group members could be subject to the benefit offset rules through September 30, 2022.

[12]   SSA will apply current rules at the end of the 60-month BOND participation period and will terminate the benefits of those engaged in SGA after any remaining GP months have been used.

[13]   The BOND sample includes disabled workers, disabled adult children, and widow(er)s receiving disabled widow(er)'s benefits.

All BOND sample members were either included in Stage 1 or were solicited for participation in Stage 2.[14] SSDI beneficiaries concurrently receiving Supplemental Security Income (SSI) benefits were randomly assigned into Stage 1, as they were not eligible for Stage 2 participation.

As would be expected if random assignment were properly implemented (and given the very large sample sizes), all differences in baseline characteristics between the T1 and C1 groups are small and appear to be attributable to chance. In Appendix B, an omnibus test for differences across all characteristics shows no statistically significant difference between the groups (Exhibit B-6). Baseline equivalence increases our confidence that any impact estimate that differs from zero at a statistically significant level represents a real impact of the interventions rather than systematic preexisting differences between the two groups or their environments.

The BOND implementation team mailed outreach letters to the T1 subjects between May and August 2011. The letters informed T1 sample members of their opportunity to use the benefit offset and to provide information about how to obtain work incentives counseling and answers to questions about the demonstration. As required by law, given the changes to SSDI rules for T1 subjects, SSA sent a follow-up letter notifying the subjects of the implications of the SSDI rule changes. The BOND implementation team sent one additional letter (2012 to 2014) and attempted to reach T1 subjects by phone (2012 to 2014) to remind them of BOND rules and to provide information about how to contact WIC and other demonstration staff, who could answer questions about the offset rules.

1.3.2.   Stage 2 Sample

Stage 2 used an experimental design to learn about the impacts of the benefit offset for beneficiaries considered most likely to use it—recruited and informed SSDI-only volunteers. Exclusion of those SSDI beneficiaries also receiving SSI (that is, concurrent beneficiaries) is based on the theoretical prediction that the SSI work incentives and the interaction between SSI and SSDI would make them less likely to use the offset. Stage 2 also examines the effects of enhanced work incentives counseling (EWIC) services relative to current law and relative to WIC services. The implementation team recruited SSDI-only beneficiaries to volunteer for Stage 2. The implementation team randomly assigned Stage 2 volunteers to one of three groups: an offset-plus-WIC group (T21 subjects), an offset-plus-EWIC group (T22 subjects), and a current-law benefits group (C2 subjects).[15] The primary difference between EWIC and WIC is that EWIC staff take a proactive approach to contacting beneficiaries on an on-going basis to inform them about the BOND demonstration, work incentives, and opportunities for employment services. WIC staff wait for beneficiaries to initiate contact before delivering services.

Of those in the Stage 2 recruitment pool, 5.32 percent volunteered.[16] The Stage 2 outreach and recruitment was intended to produce a select sample of SSDI beneficiaries, distinct from the national

---

[14]   Once beneficiaries were randomly assigned in Stage 1, they could not change their assignment or opt out of the demonstration. Subjects could choose not to participate in the survey or focus group data collection, however.

[15]   Exhibit B-7 shows the baseline characteristics of Stage 2 volunteers. All differences in baseline characteristics among the Stage 2 assignment groups are small and appear to be attributable to chance. Across all characteristics, there is no statistically significant difference between groups.

[16]   This percentage is weighted to adjust for the stratified design of the recruitment sample.

SSDI caseload in their likelihood to use the benefit offset. Differences in the 2011 employment rates confirm that the Stage 2 sample is indeed distinct from the Stage 1 sample. Altogether, 36 percent of C2 subjects were working in 2011, compared with 14 percent of C1 subjects. Given the self-selected nature of the Stage 2 sample, findings from the Stage 2 impact analysis do not generalize to the national SSDI caseload or to any easily identifiable subpopulation. Instead, they generalize to an unobservable population: SSDI-only beneficiaries in the nation who would have volunteered for BOND had they been solicited.

1.4. Organization of the Report

This report documents results for the Stage 1 and Stage 2 analyses.[17] The report presents findings from the process and participation analysis through 2016, the sixth calendar year of implementation. For the impact and benefit-cost analyses, the report presents findings through 2015, the fifth year of implementation. The report also summarizes a large set of Stage 1 and Stage 2 impact findings on self-reported beneficiary outcomes three years after random assignment. The self-reported outcomes include knowledge and understanding of offset rules, perceived barriers to employment, employment services and other services used to facilitate employment, workforce outcomes, and health outcomes. The report also synthesizes findings from the Stage 1 and Stage 2 analyses to draw cross-cutting lessons about the benefit offset. Previous reports provide interim findings about Stage 1 and Stage 2. A full listing of all BOND evaluation reports appears in Exhibit 1-4.

The remainder of the report is structured in eight chapters. Chapter 2 describes the data sources and analytic methods used in the report.

Chapter 3 describes the BOND work incentives counseling and fidelity of implementation and beneficiaries' use of work incentives counseling. The chapter compares receipt of regular work incentives counseling to enhanced counseling to assess fidelity to the design for more intensive use of enhanced counseling. Chapter 3 also analyzes how well treatment and control beneficiaries understand the relevant rules that govern how their level of earnings affects their SSDI benefit amount.

Chapter 4 analyzes beneficiary response to the offer of the benefit offset to replace the SSDI cash cliff. The chapter examines the take-up of the benefit offset, and timing of offset use over the follow-up period.

Chapter 5 presents impacts of the benefit offset on earnings, SSDI benefits, and other outcomes. Chapter 6 analyzes the costs and benefits of the offset. Chapter 7 draws lessons from analyses that cut across the components of the evaluation. The final chapter presents conclusions and a summary of key findings. Volume 2 of this report contains technical appendices that provide additional details and supporting material.

---

[17] Although this is the final report for the demonstration's evaluation period, the evaluation will also examine impacts on earnings, employment, and SSDI benefits through 2016 in two reports produced in 2018.

Exhibit 1-4.   Reports on BOND Participation, Process, and Impact Analyses for Stage 1 and Stage 2

Analysis	Stage 1 Reports	Stage 2 Reports
Participation and Process Analysis	• *Stage 1 Early Assessment Report* (Wittenburg et al. 2012)	• *Stage 2 Early Assessment Report* (Gubits et al. 2013)
	• *Process Study Report* (Derr et al. 2015)	
Impact Analysis	• *First-Year Snapshot of Earnings and Benefit Impacts for Stage 1* (Stapleton et al. 2013) • *Second-Year Snapshot of Earnings and Benefit Impacts for Stage 1* (Stapleton et al. 2014) • *Third-Year Snapshot of Earnings and Benefit Impacts for Stage 1* (Wittenburg et al. 2015) • Sixth-Year (2016) *Snapshot of Earnings and Benefits Impacts for Stage 1* (forthcoming)	• *First- and Second-Year Snapshot of Earnings and Benefit Impacts for Stage 2* (Gubits et al. 2014) • *Fourth-Year Snapshot of Earnings and Benefit Impacts for Stage 2* (Geyer et al. 2018a) • *Sixth-Year (2016) Snapshot of Earnings and Benefit Impacts for Stage 2 (forthcoming)*
Integrated Participation and Process Analysis and Impact Analysis	• *2016 Stage 1 Interim Process, Participation, and Impact Report* (Hoffman et al. 2017) • *2017 Stage 1 Interim Process, Participation, and Impact Report* (Croake et al. 2017)	• *2015 Stage 2 Interim Process, Participation, and Impact Report* (Gubits et al. 2017) • *2017 Stage 2 Interim Process, Participation, and Impact Report* (Geyer et al. 2018b)
	• *Final Evaluation Report* (this report)	

2.   Data Sources and Methodology

This chapter describes the data sources and methodology used for the process, participation, and impact analyses of the evaluation. Further technical information on the impact analysis methodology is provided in Appendix B of Volume 2 of this report. The methodology for the benefit-cost analysis is described in Appendix G of Volume 2.

2.1.   Data Sources

This section reviews the sources of data used in the evaluation. Exhibit 2-1 displays data collected from BOND demonstration operations data, BOND subjects, SSA administrative systems, and demonstration staff. The exhibit describes the data and the information they provide for the evaluation.

Exhibit 2-1.   Data Sources Used in the BOND Evaluation

Data Source	Description	Information Provided
From Demonstration Operations		
BOND Operations Data System (BODS) and Beneficiary Tracking System (BTS)	• Data management system developed for BOND. • The Beneficiary Tracking System (BTS) documents beneficiaries' contacts with the demonstration and information from SSA regarding SGA cessation and use of the benefit offset.	• Recruitment and enrollment • Random assignment result • Use of work incentives counseling • Use of offset • Benefit adjustment details (timing) *Used in process analysis, participation analysis, and benefit-cost analysis*
From BOND Subjects		
Stage 1 36-month survey **May 2014 to February 2015** **(n= 5,735, 55 percent)**	• Telephone or in-person survey (40 minutes). • Conducted a median of 39 months after random assignment.	• Employment and earnings at time of survey • Barriers to work • Receipt of education and training • Health and functional status • Income and income sources • Knowledge of SSDI benefit rules/offset • Receipt of services • Demographic information *Used to measure outcomes for the impact analysis*
Stage 2 Baseline survey **March 2011 to September 2012** **(n = 12,660, 99 percent)**	• In-person survey (50 minutes) conducted immediately before random assignment. • Completed for the full sample of families randomly assigned.	• Education and training • Current employment status • Employment history from 12 months prior to random assignment • Transportation • Barriers to employment • Health and functional status • Health insurance • Demographic information *Used to describe sample, test baseline equivalence, measure pre-BOND employment status, earnings, health status, and understanding of SSDI benefit rules. Baseline data are also used to construct covariates for the impact analysis*

Data Source	Description	Information Provided
Stage 2 12-month survey **March 2012 to January 2014** **(n = 10,713, 84 percent response)**	• Telephone or in-person survey (45 minutes). • Conducted a median of 13 months after random assignment.	• Knowledge of SSDI benefit rules/offset • Receipt of services • Employment and earnings at time of survey *Used to measure knowledge of offset and current law rules and participation in BOND work incentives counseling*
Stage 2 36-month survey **March 2014 to February 2016** **(n= 9,684, 76 percent response)[18]**	• Telephone or in-person survey (60 minutes). • Conducted a median of 39 months after random assignment.	• Employment and earnings at time of survey • Employment and earnings history from random assignment • Barriers to work • Receipt of education and training • Health and functional status • Income and income sources • Transportation • Knowledge of SSDI benefit rules/offset • Receipt of services • Demographic information *Used to measure outcomes for the impact analysis*
Telephone interviews with Stage 2 treatment subjects **August to October 2012**	• In-depth one-on-one telephone interviews with 140 Stage 2 treatment subjects.	• Experience with outreach and enrollment • Experience with BOND benefits counselors • Benefit adjustment process *Used to interpret impact findings and to evaluate BOND implementation*
Beneficiary focus groups **June to August 2012 and August to October 2013**	• Conducted focus groups with 138 Stage 2 treatment subjects in each BOND site during site visits.	• Understanding of offset • Use of BOND work incentives counseling • Opinions about BOND • Work experience, barriers to employment *Used to interpret impact findings and to evaluate BOND implementation*
Telephone Interviews with work-oriented treatment subjects **September to October 2015**	• In-depth one-on-one telephone interviews with 60 BOND T1, T21, and T22 subjects with work experience.	• Work experience and use of offset *Used to compare and contrast experiences of BOND subjects who work but do not earn enough to use the offset, who use the offset for a short period, and who use the offset for a longer period*

---

[18] Stage 2 subjects entered the survey sample at 36 months after random assignment. During the fielding period of the 36-Month Survey, the evaluation team received monthly updates from the implementation team identifying Stage 2 subjects who had died or withdrawn from the study prior to entering the survey sample. A total of 806 subjects did not enter the survey sample due to death or withdrawal from the study. Of the 12,148 subjects who did enter the survey sample, 9,830 subjects (including 146 subjects dropped from the final analysis sample—see Appendix B, Section B.2.5) completed the survey, a completion rate of 81 percent.

Data Source	Description	Information Provided
From Administrative Data Systems		
Disability Analysis File (DAF)	• Longitudinal information on SSDI beneficiaries. • The DAF is updated annually and contains information extracted from a variety of SSA source files on all SSI and SSDI beneficiaries.	• Non-countable income (such as value of impairment related work expenses) that SSA deducts from earnings to calculate benefits. *Used to measure impacts of BOND interventions on employment and earnings*
Master Earnings File (MEF)[19]	• Longitudinal information on wages and self-employment income reported to the IRS. • MEF measures Social Security Earnings, earnings that are taxable for Social Security purposes and capped at a maximum taxable amount ($118,500 in 2015).	• Annual earnings for all BOND subjects *Used to measure impacts of BOND interventions on annual earnings*
Master Beneficiary Record (MBR)	• The MBR contains information about SSDI beneficiaries' claim, payment amounts, and payee information for the benefits.	• SSDI benefits for all BOND subjects • Demographic information for the full BOND sample (date of onset, duration receiving SSDI, primary impairment) *Used to measure baseline characteristics to test equivalence, to form subgroups for the impact analysis, and to construct covariates used in the impact analysis. MBR is the source data used in the participation analysis to track use of offset.* *Used to measure impacts of BOND interventions on SSDI benefits*
Supplemental Security Record (SSR)	• SSR records monthly SSI benefits.	• SSI benefits for BOND subjects who are concurrent SSDI/SSI beneficiaries *Used to measure impacts of BOND interventions on SSI benefits*
Master Beneficiary Record, Disabled Beneficiary and Dependent (DBAD)	• Provides monthly snapshots of SSDI program activity, reflecting program activity at the time the data were pulled.	• Monthly measures of work-related overpayments *Used to analyze prevalence and impacts of the BOND interventions on work-related overpayments*
Electronic Work Reporting (eWork)	• SSA staff use the eWork tool to conduct Work Continuing Disability Reviews (work CDRs).	• Frequency and type of work CDRs for BOND subjects *Used to measure costs of administering current law SSDI rules and benefit offset rules*
Numident File	• The Numident is the master file of assigned Social Security Numbers and stores information on dates of death.	• Dates of death for deceased SSDI beneficiaries. *Used to establish Stage 1 sample and to measure incidence and timing of death*

---

[19]　Because the data are collected by the IRS and are therefore subject to IRS access rules, SSA staff have direct access to MEF data, but contractors do not. Consequently, qualified SSA staff accessed the data; submitted programs developed by the BOND Evaluation Team to estimate impacts, reviewed output to ensure that it complied with privacy requirements, and then transmitted the output to the evaluation team. The MEF earnings data are updated annually. The earnings data for this report were extracted in November 2016.

Data Source	Description	Information Provided
Case Service Report (RSA-911)	• Individual-level data from state Vocational Rehabilitation (VR) agencies maintained by the Rehabilitation Services Administration (RSA) within the U.S. Department of Education.	• Use of VR services *Used to measure impacts of BOND interventions on VR service use and to measure costs of VR service use*
From Demonstration Staff		
Telephone Focus Groups with WIC and EWIC Staff **October 2014 and July to August 2016**	• Two rounds of telephone focus groups with WIC and EWIC supervisors and counselors in the 10 BOND sites.	• Disability service environment • BOND organizational and staffing infrastructure • WIC/EWIC services • Payment problems associated with the benefit adjustment process • Preparing for the end of BOND • Influence of the offset on beneficiaries' behavior • Successes and challenges *Used to evaluate BOND implementation and assess fidelity to the design*
Site visits to BOND site offices, WIC and EWIC providers, and other service providers **April to May 2011, August to November 2011, June to October 2012, and August to October 2013**	• Three rounds of site visits to BOND sites.	• Documented implementation start up, Stage 2 recruitment • Interviewed BOND staff, observed activities *Used to evaluate implementation and BOND start up activities*
Telephone Interviews with SSA and BOND Implementation Team **September to December 2012, April 2014, January to February 2015, December 2015, July 2017**	• Interviews with Contractor team Implementing BOND. • Interviews with SSA staff responsible for BOND operations.	• Work CDR processes • Administering the offset • Collection process for Annual Earnings Estimates • BOND systems (BSAS) • Improper payments *Used to evaluate implementation*

2.2.   Methodology

This section discusses methods used to conduct the process, participation, impact and benefit-cost analyses for the BOND evaluation.

2.2.1.   Process Analysis

The BOND evaluation included a process analysis to document how SSA and the implementation team implemented the BOND interventions. The process analysis evaluates the fidelity of implementation to the original design and provides context for interpreting the BOND impact analysis. Chapter 3 (work incentives counseling), Chapter 4 (use of the benefit offset and overpayments), and Chapter 7

(interpretation and implications) of this report all present findings from the BOND process analysis. [20] The process analysis has collected data from beneficiary focus groups conducted during site visits to the BOND sites, beneficiary interviews conducted by telephone, focus group discussions with WIC and EWIC providers conducted by telephone, interviews with the SSA BOND operations team, and interviews with BOND implementation team members from Abt Associates, Mathematica Policy Research, and other implementation partners. The process analysis team also used administrative data from the BOND Operations Data System (BODS) to measure demonstration service delivery.

The process study team used several approaches to identify key themes from qualitative data. The primary approach was to code and analyze responses from beneficiary focus groups and interviews, WIC and EWIC staff focus groups and site visit interviews, and interviews with SSA staff and implementation team members. The team identified themes that emerged for all interviewees. The analysis team then compared and contrasted responses for different types of staff roles (WIC and EWIC) and for different types of BOND subjects (for example, those who did and did not use the benefit offset.)

The process study team also reviewed responses to identify details, illustrations, and other information to provide context for interpreting findings from the BOND participation and impact analyses.

In the 2014 and 2016 WIC and EWIC focus groups, in addition to open-ended discussion questions, the process study team conducted online polls to capture answers to multiple-choice questions. For each topic in the discussion, the facilitators asked staff to respond to the poll questions, waited for responses, and then began the group discussion. In that way, responses to the poll questions were not influenced by the group discussion. We use counts and percentages to describe participants' responses to polling questions. These statistics help to supplement the open-ended discussion questions by ensuring that all participants had the opportunity to respond to an identical set of questions.

2.2.2.   Participation Analysis

The BOND evaluation includes a participation analysis that describes BOND subjects' use of BOND benefits counseling services, work effort, and use of the $1 for $2 benefit offset. The participation analysis examines patterns of participation by subgroups and characteristics that distinguish offset users from other subjects. In Stage 2, the participation analysis also analyzes recruitment, to identify beneficiary characteristics that are associated with volunteering for the demonstration. Chapters 3, 4, and 7 present results from the participation analysis. The participation analysis relies on demonstration operations data, information from the beneficiary surveys, and SSA administrative data. In the participation analysis, the evaluation team used BODS data as the source for descriptive statistics on WIC and EWIC counselors' caseloads, beneficiaries' receipt of BOND counseling, the percentage of T21 and T22 subjects who completed steps toward benefit adjustment, and the duration of offset use. The analysis team used a combination of BODS data and SSA administrative records to identify beneficiary characteristics associated with offset use. SSA's Master Beneficiary Record (MBR) data tracked the percentage of T1 subjects whose benefits were adjusted according to offset rules over the follow-up period. Finally, we used SSA's Master Beneficiary Record, Disabled Beneficiary and Dependent (DBAD)

---

[20]   Appendix A describes data collection efforts for the process analysis in more detail. In addition, readers can refer to earlier reports (Derr et al. 2015, Wittenburg et al. 2012, Gubits et al. 2013, and Gubits et al. 2017*)* for more information on these data collection activities.

files to create statistics on overpayments, and a combination of BODS data and DBAD files to identify overpayments that occurred at different points in the benefit adjustment process.[21]

For this report, we analyze benefits overpayments made during the period of 2011 to 2015 for all BOND subjects. Although we are interested in overpayments beyond 2015, we limited our analysis to the pre-2016 period because of the often lengthy lag between overpayment occurrence and SSA's discovery of the overpayment. Even having limited the analysis to the pre-2016 period, because SSA may continue to identify new overpayments as it receives and processes information, the statistics we present are lower-bound estimates of the prevalence of overpayments.

2.2.3. Impact Analysis

The central objective of the BOND evaluation is to estimate the impact (or the effect) of the benefit offset policy on beneficiaries' earnings and SSDI benefit receipt, relative to the current-law SSDI rules. This report presents estimates of the impacts of the Stage 1 treatment and of the two Stage 2 treatments (offset-plus-WIC and offset-plus-EWIC) on outcomes measured in administrative and survey data. The Stage 2 impact analysis also examines the impact of enhancements to work incentive counseling (EWIC versus WIC). Chapter 5 presents impact findings for Stages 1 and 2. Chapter 7 discusses cross-cutting lessons from the Stage 1 and Stage 2 impact analyses. This section describes the analytic methods used to estimate the impacts.

Confirmatory Outcomes
The key outcomes of the BOND evaluation are total earnings and total SSDI benefits during the follow-up period.

We designate these outcomes as "confirmatory" to show that they have been pre-specified as the two most important outcomes among the many measured by the evaluation. All other outcomes in the study are considered "exploratory" (i.e., not confirmatory). Because we have two key outcomes rather than one, the probability of a "false positive" finding—i.e., appearing to find an effect, when there is not truly an effect—is higher than if the study only had a single key outcome. In order to address this higher probability of a chance finding of a statistically significant impact on a confirmatory outcome, we use a higher standard of evidence when determining the statistical significance of impacts on these outcomes (see below).

The source of the earnings data is SSA's Master Earnings File (MEF). The data available for this evaluation measure earnings that are taxable for Social Security purposes.[22] For the Stage 1 impact

---

[21] Appendix E presents additional details about construction of the overpayment measure.

[22] The earnings measure is subject to two limitations. First, about 6 percent of people in the U.S. work force hold jobs not covered by Social Security taxes. Second, taxable earnings for Social Security are capped at a maximum amount (e.g., $118,500 for 2015). Of the two limitations, we do not expect the cap to be a problem for the analysis because very few study subjects have earnings at or above the maximum taxable amount. In 2015, 0.04 percent of Stage 1 subjects had earnings equal to or above the 2015 maximum taxable amount. In addition, beneficiaries who are earning at or above that amount are unlikely to have a behavioral response to the offset. Non-covered jobs constitute a larger omission. It is not feasible for this evaluation to obtain a more comprehensive measure of earnings from administrative data. As a result, reported findings for earnings, employment, and the proportion with earnings above BYA have a small downward bias. In addition, the

analysis, we use a cumulative earnings measure of total earnings from 2011 to 2015.[23] For Stage 2, the earnings measure is total earnings from 2012 to 2015.[24] The Stage 2 follow-up period begins in 2012 rather than 2011 because enrollment of volunteers into the study finished in 2012. Total earnings amounts for the follow-up period are adjusted for inflation and presented in 2016 dollars.

The source for the SSDI benefits measure is SSA's Master Beneficiary Record (MBR). For Stage 1, we measure total SSDI benefits from May 2011 (the first month after random assignment) to December 2015. For Stage 2, the follow-up period is 2012 to 2015. This measure incorporates all retroactive adjustments made through the data extraction month: May 2017. Trust Fund expenditures for benefits during this period may differ somewhat from this measure due to retroactive adjustments after May 2017 and unrecovered overpayments. The measure used in this report differs from the benefits measure used in earlier reports, for which sufficient time had not elapsed for retroactive adjustments to occur. See Appendix A for a comparison of this benefits measure to the benefits measure used in earlier reports. As with earnings, the dollar amounts for SSDI benefits are adjusted for inflation and presented in 2016 dollars.

Exhibit 2-2 shows the definitions of the Stage 1 and Stage 2 confirmatory outcomes.

Exhibit 2-2.   Confirmatory Outcomes for BOND Evaluation Impact Analysis

Confirmatory Outcome	BOND Stage	
	Stage 1	**Stage 2**
Earnings	2011 to 2015 total earnings	2012 to 2015 total earnings
SSDI benefits	May 2011 to December 2015 total benefits due (as recorded in May 2017)	January 2012 to December 2015 total benefits due (as recorded in May 2017)

Note: All dollar amounts are inflation adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).

Stage 1 Impact Estimation Methodology

The goal of Stage 1 of the demonstration is to make inferences about the impact of the benefit offset if it had been applied to all SSDI beneficiaries in the nation who met the BOND eligibility criteria as of May 2011. The statistical design of the demonstration supports the production of unbiased impact estimates and their standard errors for that nationwide population. The estimation of the standard errors accounts for

---

estimate of the impacts of the offset on earnings, employment, and proportion working above BYA may have a small downward bias if some who are encouraged to work choose jobs not covered by Social Security.

[23]   The earnings data from the Master Earnings File capture earnings by calendar year, preventing precise alignment of the earnings outcome to the Stage 1 random assignment date of May 2011.

[24]   The annual earnings data do not align with the study enrollment dates of Stage 2 volunteers, which stretch from March 2011 to September 2012. For the 40 percent of volunteers who enrolled in 2011, the entire calendar year of 2012 was post-random assignment. For the other 60 percent of volunteers, less than the entire year of 2012 was post-random assignment. For the entire Stage 2 sample, 77 percent of person-days in 2012 were post-random assignment.

random variation associated with both the selection of the BOND sites and the assignment of subjects in those sites to the T1 and C1 groups.

To estimate impacts on outcomes observed in administrative data, we compare the mean of a given outcome (for example, total earnings for 2011 to 2015) for the T1 group to the mean of the same outcome for the C1 group. The sample means are weighted for differences in (1) site-selection probabilities, and (2) sampling rates into T1 and C1 status across sampling strata. Survey outcome means are additionally weighted for differences in survey response propensities, in order to address the possibility of non-response bias.

For both types of outcomes, using (weighted) linear regression methods, the means are also adjusted for the effects of small random differences in baseline characteristics.[25] The adjustments for differences in baseline characteristics also reduce the standard errors of the impact estimates.

For each outcome, we test the null hypothesis of no impact. Each test uses a specified level of statistical significance. For example, a 10 percent significance level means that, if the null hypothesis is true, then there is a 10 percent chance that the test will mistakenly reject it.

When discussing the impact estimates, we use particular language to signify differing levels of confidence that a non-zero impact has occurred. When the null hypothesis of no effect can be rejected with 99 percent confidence (that is, with 0.01 statistical significance), we state that the estimate *provides strong evidence* that the benefit offset had an effect on the tested outcome. When the null hypothesis of no effect can be rejected with 95 percent confidence (that is, with 0.05 statistical significance) but not 99 percent confidence, we state that the estimate *provides evidence* that the offset had an effect on the tested outcome. Finally, when the null hypothesis of no effect can be rejected with 90 percent confidence (that is, with 0.10 statistical significance) but not 95 percent confidence, we state that the estimate *provides some evidence* that the offset had an effect on the tested outcome.

All impact estimates are "intent to treat" estimates. They capture the mean impact of applying the BOND offset rules to the earnings of *all* T1 subjects, regardless of how many subjects work or use the offset. Hence, our average impact measures reflect no impacts on T1 subjects who do not respond to the offset and whose earnings or benefits are not affected by it. We chose to generate "intent to treat" estimates because of a strong policy interest in understanding the BOND offset's effects on all SSDI beneficiaries as opposed to (for example) on only those beneficiaries who use the offset.

We make a multiple-comparison adjustment for the tests of impacts on the two confirmatory outcomes. The adjustment is needed because we are testing more than one outcome, thereby making the probability of a Type I error (rejecting the null hypotheses if it is true) larger than the significance level for an individual test. To compensate, we adjust the test statistics for each of the two confirmatory outcomes so

---

[25]   Appendix B in Volume 2 of this report provides a full description of the estimation method and the construction of analysis weights.

that the probability of rejecting the joint null hypothesis of no impact on both confirmatory outcomes in Stage 1 is equal to the specified significance level if the null hypothesis is true.[26]

We make no multiple-comparison adjustment to the tests for exploratory outcomes (i.e., all outcomes except earnings and SSDI benefits). Readers are advised to give less evidentiary weight to any individually significant result from an exploratory test than they would to an equally significant result from a confirmatory test.

Stage 2 Impact Estimation Methodology

For Stage 2, the impact analysis compares mean outcomes for the T21, T22 and C2 groups in three pairwise comparisons:

> (1) T21 vs C2;
> (2) T22 vs C2; and
> (3) T21 vs T22.

For outcomes derived from administrative data, the sample means are weighted for differences in site-selection probabilities and differences in sampling rates into the Stage 2 solicitation pool across sampling strata. For outcomes derived from survey data, the sample means are additionally weighted for survey response propensities, in order to address the possibility of non-response bias. For both these types of outcomes, using (weighted) linear regression methods, the means are adjusted for the effects of small random differences in baseline characteristics. The adjustments for differences in baseline characteristics also serve to reduce the standard errors.[27]

The Stage 2 impact analysis has a total of six confirmatory hypothesis tests: tests of impacts on the two confirmatory outcomes in each of the three pairwise comparisons. We group the four tests in the T21 vs. C2 and T22 vs. C2 comparisons together because these two comparisons involve impacts of the benefit offset policy versus current law. We perform a multiple comparison procedure on these four tests together to adjust the p-values of the tests. We perform a separate multiple comparison procedure to adjust the p-values of the two confirmatory tests in the T22 vs. T21 (EWIC vs. WIC) comparison. As in the Stage 1 impact analysis, we make no multiple comparison adjustment to the tests for exploratory outcomes.[28]

---

[26] Our approach adjusts the *p*-values for the confirmatory outcomes by using the Westfall-Young stepdown method. Appendix B presents details of the *p*-value adjustments for tests of impacts on the confirmatory outcomes. See Schochet (2009) for further discussion of the multiple-comparisons problem.

[27] Because of the smaller size of the Stage 2 sample, the Stage 2 estimation method is able to omit some steps used in the Stage 1 estimation method to speed computations. Details of the Stage 2 estimation method are provided in Appendix B of Volume 2 of this report.

[28] We note that the multiple comparisons problem is addressed separately in Stages 1 and 2, with each stage having its own adjustment procedure. This is appropriate as the stages draw inferences about distinct reference populations.

Subgroup Analysis

We estimate impacts on outcomes observed in administrative data for the full Stage 1 and Stage 2 assignment groups and for several pairs of subgroups. We treat all subgroup analyses, including the subgroup analyses for the confirmatory outcomes (i.e., earnings and SSDI benefits), as exploratory.

For Stage 1, we estimate impacts for seven pairs of beneficiary subgroups. The impact estimation method we use for each subgroup mirrors the impact estimation method we use for the entire sample. Specifically, we use t-tests to examine whether impact differences between subgroups are statistically significant.

The first subgroup pair is defined by duration of SSDI benefit receipt at the point of solicitation into the demonstration.[29] The duration subgroups are of interest. Earlier research (Liu and Stapleton 2011) and program rules suggest that subjects who have been on the rolls for a *short duration* (defined here as three years or less at baseline) may respond to the benefit offset differently from those who have been on the rolls for a *long duration* (more than three years). More specifically, we expect more short-duration subjects to work than long-duration subjects. However, we expect that it will take longer for short-duration subjects to see their benefits adjusted because, unlike long-duration subjects, they will have completed fewer TWP and GP months at the outset of the demonstration.

The second subgroup pair divides the sample by SSI payment receipt status at baseline. Relative to SSDI beneficiaries who do not receive SSI payments, concurrent beneficiaries—those who receive SSI and SSDI benefits at the same time—have less income and fewer assets and are more likely to be Medicaid enrollees. These differences may create different barriers to employment for the two subgroups. In addition, the work incentives for SSI differ from the work incentives for SSDI, with SSI recipients experiencing a $1 for $2 payment offset for earnings above a $65 monthly earnings disregard and a $1 for $1 payment reduction for unearned income above a monthly unearned income disregard. As described further in the *Evaluation Analysis Plan* (Section 2.1.2), the interaction of the two work incentives in the context of current law reduces the value of the SSDI benefit offset for concurrent subjects relative to SSDI-only beneficiaries with the same SSDI benefit amount. Hence, the expectation is that concurrent beneficiaries will be less responsive to the benefit offset than otherwise comparable SSDI-only beneficiaries.

Other subgroup pairs are defined by (1) employment status in 2010 before entry into the demonstration, (2) whether the participant lives in a state with a Medicaid Buy-In (MBI) program, and (3) age at baseline. We expect that subjects who are employed or who are younger at baseline will be more likely to use the benefit offset because they face higher opportunity costs of not working. For example, those who worked in 2010 may be able to increase earnings enough to take advantage of the offset more readily than beneficiaries who are not already working. Compared to older beneficiaries, younger beneficiaries may also gain more economically by changing fields through job training or other means because they have more years before retirement to gain earnings by investing in a new career.

Most states now offer an MBI program for people with disabilities who may be concerned that they will lose their Medicaid coverage if they enter or return to the workforce. Commercial or employer-based

---

[29]   We used the disability adjudication date as the start date for receiving SSDI. When this date was missing, we used the date of initial SSDI entitlement.

health insurance might not provide coverage for services and supports that enable people with disabilities to work and live independently. Therefore, theory predicts that study subjects with access to an MBI program will be more likely to use the benefit offset than study subjects without MBI access because they face a lower risk of losing health insurance when their earnings change. [30]

The remaining two subgroup pairs are defined by specific disabilities: a primary impairment of major affective disorder and a primary impairment of back disorder, both at baseline. The incidence of these two primary impairments has grown significantly in recent years. It is therefore of interest whether the earnings and benefits of the two affected groups are more or less sensitive to the introduction of the benefit offset relative to those of beneficiaries with other impairments.

The Stage 2 subgroup analysis examines five of these seven subgroup pairs. The Stage 2 analysis omits the concurrent receipt of SSI status subgroup pair because concurrent beneficiaries were not solicited for enrollment in Stage 2. It also omits employment in 2010 as a subgroup-defining characteristic, and instead uses the employment status at study enrollment, collected on the Stage 2 baseline survey. Lastly, we include a seventh subgroup pair in the Stage 2 subgroup analysis defined by educational status at baseline (also collected on the baseline survey). Beneficiaries with higher education may have more employment options than those with lower levels of education, in part because higher education may give more options for changing fields—for example, from construction to information systems. They may also be better able to understand the offset rules and, therefore, be more likely to change their behavior in response.

A finding that impacts differ across two subgroups does not necessarily imply that the variables used to define the two subgroups *caused* the difference. This point is especially important for the MBI subgroups, as access to an MBI program may be correlated with other features of the policy or economic environment that also affect impacts. If the impacts between two subgroups do not differ in a statistically significant manner, we consider findings for the full sample to be the best available evidence on each individual subgroup. This practice is adopted because the full sample yields more precise findings (that is, has smaller standard errors) than the smaller subgroup samples (Bloom and Michalopoulos 2013).

The next chapter reports on analyses of BOND work incentives counseling, receipt of service external to BOND, and BOND subjects' knowledge of how earnings affect SSDI benefits.

---

[30]   We defined access to the Medicaid Buy-In based on state of residence just before random assignment. We categorized beneficiaries residing in Alabama, Colorado, Washington, DC, and Florida at that time as not having access to the Medicaid Buy-In. Beneficiaries in the remaining states did have Medicaid Buy-In access.

3.   BOND Work Incentives Counseling, Outside Services, and Knowledge of How Earnings Affect Calculation of Benefits

Many beneficiaries have trouble understanding current-law SSDI work incentive rules. To help beneficiaries understand the rules, SSA funds 103 Work Incentives, Planning, and Assistance (WIPA) grantees. Counselors called Certified Work Incentives Coordinators (CWICs) working at WIPA programs offer SSDI beneficiaries information about how SSDI benefits work. In particular, CWICs help beneficiaries to understand how earnings affect benefits, and how SSDI interacts with other federal and state benefits. For beneficiaries who are interested, CWICs also may provide advice on how to seek and maintain employment or how to find employment services.

The BOND treatment groups are subject to the benefit offset rules, an alternative set of work incentives. These rules are no less complicated than current-law SSDI rules. In place of WIPA counseling, BOND offered treatment subjects work incentives counseling tailored to BOND rules. Section 3.1 of this chapter explains the design of work incentive counseling in BOND. Section 3.2 describes the delivery of that counseling and its fidelity to the BOND design. Section 3.3 examines receipt of employment services and Section 3.4 describes BOND subjects' knowledge of how earnings affect SSDI benefits. Section 3.5 summarizes the chapter findings.

3.1.   Design of BOND Counseling

This section summarizes the three types of work incentives counseling that were part of BOND: WIPA, WIC, and EWIC.

Like all SSDI beneficiaries not participating in BOND, the Stage 1 control group (C1) and Stage 2 volunteers assigned to the control group (C2) are eligible to receive work incentives counseling from a CWIC. CWICs explain the potential effects of work on SSDI and other benefits and help beneficiaries to make informed choices about work.

When a beneficiary first contacts a WIPA grantee, a CWIC begins by offering "information and referral" (I&R) services. Information and referral services consist of answering basic questions about types of benefits or work supports and determining if the beneficiary needs more individualized, ongoing support. Beneficiaries who request more individualized, in-depth services can enroll in WIPA to work with a CWIC on an ongoing basis. After enrolling in WIPA, the beneficiary works with the CWIC to develop and carry out long-term plans to use SSA work incentives and other employment supports. CWICs may also refer WIPA enrollees to employment support programs, such as state Vocational Rehabilitation (VR) agencies or Employment Networks (ENs) through the Ticket to Work program (TTW).

Stage 1 treatment subjects (T1) and Stage 2 subjects in the T21 group are eligible to receive basic work incentives counseling (WIC). WIC was designed to mimic WIPA in the type and intensity of services provided. WIC is provided to T1 and T21 subjects by counselors who have the same certification as CWICs. The only intended difference is in the content, with WIC describing the benefit offset work incentives rather than current-law work incentives.

Stage 2 volunteers assigned to the T22 group receive Enhanced Work Incentives Counseling (EWIC). The primary difference between EWIC and WIC is that EWIC staff take a proactive approach to contacting beneficiaries on an on-going basis to inform them about the BOND demonstration, work

incentives, and opportunities for employment services. EWIC staff were instructed to contact all T22 beneficiaries within two weeks of random assignment and contact them thereafter at least once per month. The requirements for contact by EWIC staff were modified in early 2014 after all T22 subjects had received at least 18 months of monthly contact. From that date forward, EWIC staff were to contact all engaged[31] T22 subjects at least quarterly, with monthly contacts for those deemed likely to use the offset.

Compared to WIC, EWIC also includes additional services. The enhancements in EWIC include a detailed employment support plan based on assessments of vocational skills and interests, and assistance to obtain the resources and support beneficiaries need to find employment, as well as the ongoing support they need to keep it. WIC staff are not supposed to conduct the assessments of vocational skills and interests or develop the employment support plan that are both part of the EWIC design, though WIC providers might refer beneficiaries to other providers to receive those types of services. To avoid cross-over treatment, counselors that provide EWIC to T22 subjects did not also provide counseling to T1 or T21 treatment subjects.

Exhibit 3-1.　　Comparison of EWIC and WIC Services

	WIC	EWIC
	Provided to T1 and T21 (Intended to be identical to WIPA)	**Provided to T22**
Outreach and engagement	Only respond to beneficiary-initiated contact; do not contact beneficiaries.	Contact beneficiaries once per month for the first 18 months after random assignment, monthly thereafter if expected to use the offset, and quarterly if not expected to use the offset.
Work Focus: (1) Barriers and needs assessment	None.	Administer psycho-social needs assessment to identify employment barriers and needs, such as transportation, skill deficits, and childcare.
Work-Focus: (2) Skills assessment	None.	Administer assessment to assess aptitude, skill; and administer a separate assessment to match skills with occupational requirements, providing average wage data to help beneficiaries evaluate earnings potential.
Developing an Employment Services Plan	Develop Benefits Summary and Analysis (BS&A) to analyze benefits and work incentives. Develop a Work Incentives Plan (WIP) that documents vocational goals, contains referral information, and describes how benefits would respond to changes in earnings.	In addition to the BS&A and the WIP, develop an Employment Services Plan (ESP) that documents vocational goals and specific plans to achieve those goals, overcoming employment barriers. Specific plans include planned referrals to VR or ENs for vocational assessments, employment planning and support.
Service Coordination	Refer beneficiary to employment services, but do not monitor service receipt or completion.	Refer the beneficiary to employment assistance services documented in the ESP. These employment services include pre-employment skills development, job search assistance, and job placement. EWICs check with beneficiaries regularly to check progress to completion.

---

[31]　An EWIC can designate a T22 subject as unengaged if the beneficiary is incarcerated, asks not to be contacted, is not responsive to repeated contact attempts, or if the beneficiary reports not being interested in employment. The BOND implementation team contacted unengaged beneficiaries twice per year to remind them of their BOND treatment status and the availability of EWIC services.

3.2.   Delivery of Work Incentives Counseling

The process and participation analyses show that counseling use is slightly higher in the T1 treatment group than in the national SSDI population. Approximately 1.1 percent of the national SSDI beneficiary population meeting BOND eligibility criteria receives WIPA counseling in a typical year. [32] The percentage of T1 subjects receiving WIC counseling grew from 2.0 percent in 2012 to a maximum of 2.9 percent in 2013, falling to 1.2 percent by 2016 (Exhibit 3-2).

A demonstration-related circumstance likely explains why T1 subjects use more counseling than do non-BOND SSDI beneficiaries: the demonstration included active BOND outreach to inform T1 subjects of their new benefit rules and the availability of benefits counseling (see Section 1.3.1).

In Stage 2, WIC staff engaged with 28 percent of T21 subjects in 2012, decreasing to 2 percent in 2016. Higher WIC take up for the T21 group compared to the T1 group might be explained by the selection of T21 subjects. Recruited and informed volunteers would be expected to be more engaged with the counseling intervention compared to T1 subjects who did not volunteer for the demonstration.

In contrast, the BOND design called for T22 subjects to receive more counseling than current-law SSDI beneficiaries. The process and participation analyses show that the differences between services that WIC and EWIC counselors delivered to Stage 2 subjects are substantial and in the expected direction.

- Compared to beneficiaries eligible for WIC (T1 and T21 subjects), beneficiaries eligible for EWIC (T22 subjects) were much more likely to have had contact with a benefits counselor.

- Beneficiaries eligible for EWIC are also more likely to receive counseling beyond basic information and referral compared to beneficiaries eligible for WIC.

- Exhibit 3-2 shows that EWIC counselor contacts peaked at 90 percent in the year Stage 2 was completed (2012) and declined over time (as designed). Still, even in the last year of observation (2016), EWIC-eligible beneficiaries were 28 percentage points more likely to have been contacted by a counselor, with 29 percent of all EWIC-eligible beneficiaries receiving individualized counseling beyond basic information and referral. To simplify the main document, certain detailed exhibits are presented in Volume II – Technical Appendices. : Exhibits D-1 and D-2 in Volume 2 present the estimated impacts of EWIC on counseling receipt.

- For the most part, EWIC counselors' use of EWIC-specific counseling tools met the benchmark rates set in the design of BOND (Exhibit D-3).

- In addition to EWIC-specific counseling tools, receipt of typical WIPA services provided by EWIC staff were high among T22 subjects. For example, 66 percent of T22 subjects received referrals to employment support services (Exhibit D-3) and 54 percent of T22 subjects worked with EWIC counselors to receive a Benefits Summary and Analysis (BS&A), which summarizes an individual's current benefits and offers individual-specific information on how the offset and other work incentives would affect the beneficiary's SSDI and other possible benefits, such as

---

[32]   SSDI beneficiaries' receipt of WIPA counseling is estimated based on published tabulations of national data on the WIPA program. (Schimmel et al. 2013).

Supplemental Nutrition Assistance Program (SNAP) benefits and health care coverage (Geyer et al. 2018b).

Exhibit 3-2.   WIC and EWIC Delivery Over Time, by Treatment Group
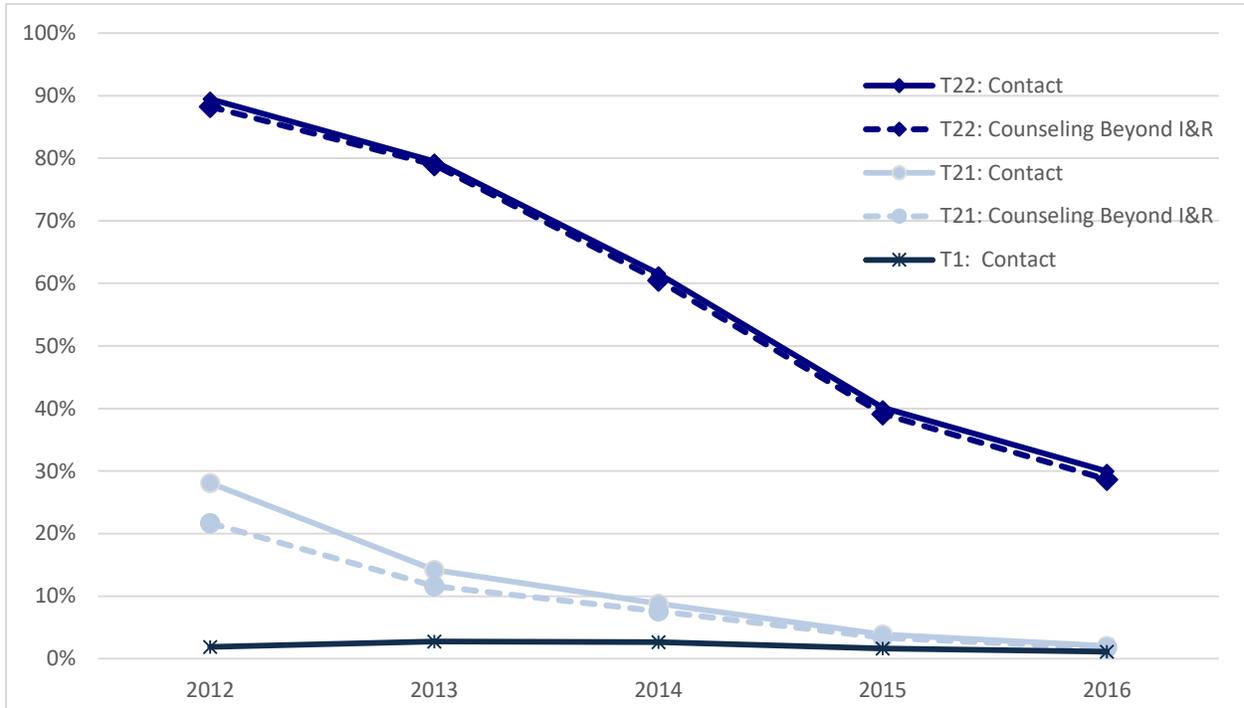
Source: SSA administrative records from the BTS and MBR and the Stage 2 Baseline Survey.

Notes: T22 and T21 means are regression-adjusted for baseline characteristics. Weight are used for T22 and T21 to ensure that the BOND subjects who met analysis criteria are representative of the national population of SSDI-only beneficiaries who would volunteer for study enrollment. Weights are used for T1 to ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment, but not regression-adjusted because comparable data on counseling receipt for C1 subjects are not available. Data on counseling beyond I&R are not available for Stage 1.

Unweighted sample sizes: T1 = 77,115, T21 = 4,854 and T22 = 3,041.

3.3.   Receipt of and Unmet Need for Employment Services Support

In addition to counseling about SSDI benefits and work incentives, BOND treatment and control subjects may seek services to prepare for, find, and maintain employment. CWICs and WIC staff do not provide such services directly. Instead, they refer beneficiaries to partner agencies such as state Vocational Rehabilitation (VR) agencies or Employment Networks (ENs) under the Ticket to Work Program (TTW). Relative to WIPA, WIC is not expected to increase use of VR or EN programs. However, in order to take advantage of the benefit offset, treatment group beneficiaries might be more likely to seek out employment services.

Despite this hypothesis, there is no evidence that T1 subjects (who were eligible to receive WIC) received services from state VR agencies or ENs at a different rate than C1 subjects (who were eligible to receive WIPA services) (Exhibit D-4 and D-5). Nor is there evidence that T21 subjects (assigned to WIC) received services from state VR agencies or ENs at different rates than C2 subjects (who were eligible to receive WIPA) (Exhibit D-6 and D-7).

In contrast to CWICs and WIC counselors, EWIC counselors not only make referrals but also act as case coordinators. For example, after referrals to services, EWIC counselors were to follow up with a beneficiary to ensure that the beneficiary is linked to services. Consistent with this program model, EWIC increased the proportion of beneficiaries receiving VR services by 3 percentage points compared to WIC, from 12 percent in T21 to 15 percent in T22 (Exhibit D-6). This is a relative increase of 25 percent of the T21 level.

Similar to its impact on VR service receipt, compared to WIC, EWIC increased the proportion of beneficiaries using an EN by 6 percentage points, from 19 percent in T21 to 26 percent in T22 (Exhibit D-6).[33] This is a relative increase of 32 percent of the T21 level. Just as for VR, most of this impact is likely due to intensive case coordination rather than the benefit offset rules alone, because the impact of EWIC relative to WIC was large while the impact of T21 relative to C2 (or T1 relative to C1) was not statistically significant.

There is some qualitative evidence that the use of state VR agencies and EN services could have been higher if it were not for the inconsistent availability of employment service providers. WIC and EWIC counselors reported long waiting lists at state VR agencies in some sites and few ENs in some areas. During the 2016 focus group discussions, counselors in three of the eight focus groups reported that they do not typically refer beneficiaries to employment support services because their recent experience suggests a long wait time and that beneficiaries will lose interest while waiting. However, in four focus groups, a few participants indicated that many beneficiaries in their caseloads do not need state VR agencies or EN services.

Consistent with this qualitative evidence of limited availability, both Stage 1 and Stage 2 treatment and control subjects indicated (in surveys) that they had unmet need for such services (Hoffman et al. 2017

---

[33]   These results are rounded.

and Geyer et al. 2018b).[34] After three years of study participation, 37 percent of Stage 1 beneficiaries and 48 percent of Stage 2 beneficiaries stated that they needed an employment support but did not receive it. The unmet need levels may be higher for Stage 2 than for Stage 1 because their perception of unmet need might be heightened by the experience of searching for a job (an experience more common among BOND study volunteers). These levels of unmet need are much higher than the percentages of T1 and T21 subjects who reached out to a WIC counselor. Perhaps some beneficiaries stated a need on the survey but were not actively trying to find employment support. Alternatively, some beneficiaries may not have known that WIC counselors can refer them to the support they need, or some may have tried WIC services but found referrals were not fruitful. At least one of these reasons leading to unmet need was partially addressed by EWIC, as EWIC staff conducted outreach.

Even among T22 beneficiaries, EWIC counseling did not alleviate unmet need. The four most common unmet needs to overcome an employment barrier were training to learn a new job or skill (23 percent), help to find a job (22 percent); on-the-job training, coaching or support services (18 percent); and transportation assistance (17 percent). Employment-related services could help address these barriers. In fact, EWIC counselors report that they provided employment-related referrals to 66 percent of the T22 beneficiaries (Exhibit D-3). However, based on beneficiary self-reported survey responses, there is no evidence that subjects eligible for EWIC had fewer or greater unmet needs than subjects eligible for WIC (Geyer et al. 2018b). Thus, for T22 subjects, either beneficiaries simply did not follow through on those referrals, or service availability was too constrained.

3.4.   Knowledge of How Earnings Affect Calculation of Benefits

In order for treatment subjects to change their work behavior relative to their work behavior under current law, they need to understand that the offset makes work more attractive. Survey data show that many Stage 1 and Stage 2 treatment subjects have limited understanding of the offset rules. Approximately three years after the implementation team sent initial mailings to the Stage 1 treatment group, only 35 percent of its members reported having heard of BOND (Hoffman et al. 2017). Given that, it is not surprising that only 29 percent of T1 subjects gave survey responses consistent with a correct understanding of how earnings affect benefits (Exhibit 3-3). The fact that only 29 percent of T1 subjects understood how earnings affect benefits under the offset rules raises concern that the Stage 1 impacts of the benefit offset on earnings, benefits, and other outcomes may be muted.

In Stage 2, all beneficiaries have heard of BOND because they were recruited for the study and provided signed informed consent. Three years after random assignment, almost all Stage 2 volunteers (96 percent) stated that they had heard of BOND (Geyer et al. 2018b). However, this outreach effort did not result in high levels of understanding of the benefit offset rules. Roughly half (48 percent) of T21 subjects correctly understood how earnings above BYA affect benefits. Compared to WIC, EWIC increased correct understanding by 4 percentage points, from 48 percentage points among T21 subjects to 52 percentage points among T22 subjects (Exhibits 3-3, 3-5).

---

[34]   In the Stage 1 control group, the four most common unmet needs to overcome an employment barrier are transportation assistance (12 percent), training to learn a new job or skill (11 percent), assistive devices (11 percent), and help to find a job (10 percent) (Hoffman et al. 2017).

Although current-law rules have not changed meaningfully in decades,[35] BOND survey data show that there is also substantial confusion among beneficiaries about current-law rules. In both stages, about half of the control group have a basic understanding of how their earnings affect their benefits under current law (54 percent of both C1 and C2 groups). Given the expectation that the Stage 2 volunteers have a greater interest in work than the Stage 1 subjects, the fact that this percentage is essentially the same in Stage 1 and Stage 2 is unexpected.

The confusion among control subjects about current-law rules may also be noted in the shares that provided the incorrect response that above-SGA earnings would lead to partial benefits. This response is consistent with the offset rules, rather than with current law. Among the C1 group, 22 percent gave this response (Exhibit 3-4). In Stage 2, this incorrect response was even more common, at 37 percent of the C2 group (Exhibit 3-5). It seems plausible that the higher rate of this response among the C2 group may be due to their greater contact with the demonstration during the recruiting and informed consent process.

One interpretation of these findings might be that most beneficiaries have no interest in working and thus pay little attention to how benefits would change with earnings. If this interpretation were correct, we would expect substantially better understanding among those working at baseline. However, we find no such evidence of differential understanding in Stage 1. In addition, Stage 2 treatment subjects who were working at baseline were not more likely to correctly understand the offset rules (Exhibit D-8). Only for the Stage 2 control group was there evidence of differential understanding. The Stage 2 control group subjects employed at baseline were more likely to report correct understanding than those not employed at baseline (62 versus 50 percent, Exhibit D-8). EWIC did not make a statistically significant difference in correct understanding between the employed at baseline and not employed at baseline subgroups.

---

[35]   The TWP was added in 1960, in which a beneficiary would work for nine months (which need not be consecutive) and earn as much as they were able without losing cash benefits. Beginning in 1992, the 9-month trial work period was changed to a rolling 9 months in any 60-month period. In 1980, a 15-month extended period of eligibility, a 36-month period of extended Medicare, and impairment-related work expenses were added to the DI program. In 1988, the extended period of eligibility was increased to 36 months. See SSA "Changes in Program Policy Influencing Program Size" for more detail (SSA 2017c).

Exhibit 3-3.   Correct Understanding of How Earnings Affect Benefits at 36 months after Random Assignment, by Assignment Group
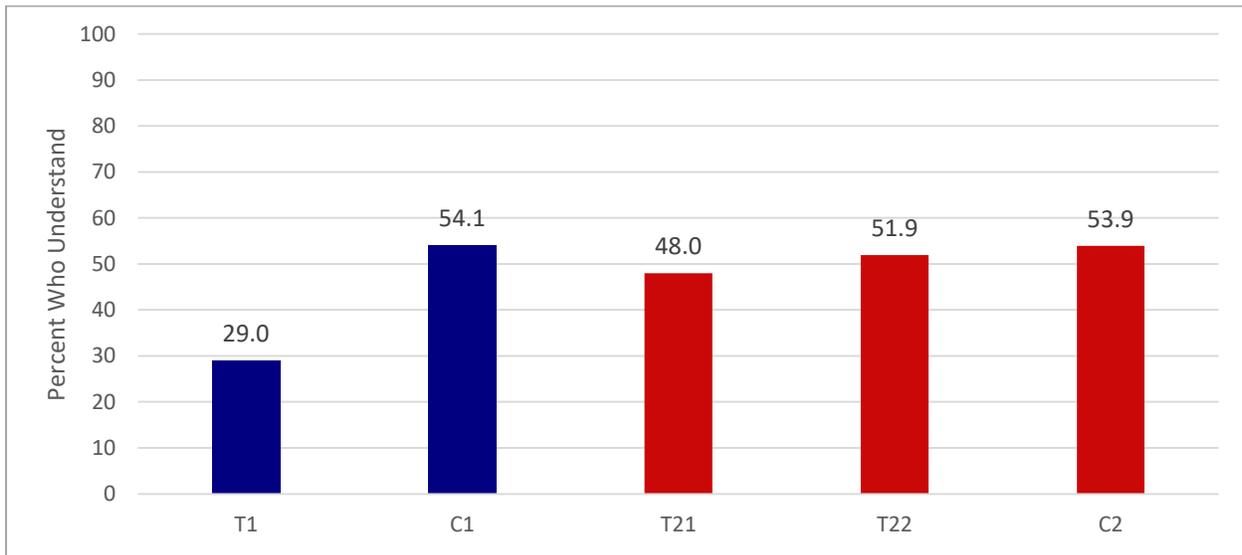
Source: BOND Stage 1 and Stage 2 36-Month Surveys.

NOTE: Based on their survey responses, we identified beneficiaries who demonstrated an understanding of benefit adjustment consistent with the rules that apply to them. The 36-month survey asked respondents how their monthly disability cash benefits would change if they were to earn more than the SGA limit after the Trial Work Period (TWP). We categorized as 'demonstrating an understanding' in the T1, T21, and T22 groups those beneficiaries whose response indicated that benefits would be reduced but not to $0 if earnings above SGA are sustained after the TWP months. We categorized as 'demonstrating an understanding' in the C1 and C2 groups those beneficiaries whose response indicated that benefits would be reduced to $0 if earnings above SGA are sustained after the TWP months.

Unweighted sample sizes: T1 = 2,916; C1 = 2,819 T21 = 3,785; T22 = 2,384; C2 = 3,661.

Exhibit 3-4.   Estimated Impacts on Stage 1 Subjects' Understanding of How Earnings Affect Benefits (at 36 Months after random assignment; correct answer in bold)

Outcome	T1 Mean	C1 Mean	Impact Estimate
If earnings are above SGA-level after TWP months, subjects who think			
Benefits would stay the same (%)	3.4	4.9	-1.5 (0.9)
Benefits would be reduced but not to $0 (%)	**29.0**	22.4	6.6*** (2.0)
Benefits would be reduced to $0 (%)	53.1	**54.1**	-1.0 (2.0)
Benefits would neither stay same, nor be reduced (%)	2.8	4.2	-1.4* (0.7)
Don't know whether benefits would change (%)	11.7	14.4	-2.7 (1.9)

Source: BOND Stage 1 36-Month Survey.

Note: The correct answers expected from T1 and C1 subjects are indicated in bold. For example, "benefits would be reduced but not to $0" is the correct answer expected from T1 subjects (Appendix Section A.4.1). Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment. The weights, however, do not account for the disproportionately low sampling rate of subjects residing in multi-subject households, especially in the T1 sample (Appendix Section B.3.4). Means and impact estimates are regression adjusted for baseline characteristics that include an indicator for whether a subject resided in a multi-subject household at baseline. Standard errors are in parentheses.

Unweighted sample sizes: T1 = 2,916, C1 = 2,819

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test with 9 degrees of freedom (resulting from a research design involving 10 study sites) and with no multiple- comparisons adjustment.

Exhibit 3-5.   Estimated Impacts on Stage 2 Subjects' Understanding of How Earnings Affect Benefits (at 36 Months after random assignment)

Outcome	Percent of T21 subjects who gave response (1)	Percent of T22 subjects who gave response (2)	Percent of C2 subjects who gave response (3)	Estimated Impact of Offset and WIC vs. Current Law (T21 vs. C2) (4)	Estimated Impact of Offset and EWIC vs. Current Law (T22 vs. C2) (5)	Estimated Impact of EWIC instead of WIC Given Offset (T22 vs. T21) (6)
If earnings above SGA-level beyond TWP months, percent of all subjects who think:						
Benefits would stay the same	1.8	2.4	1.9	-0.1 (0.4)	0.4 (0.6)	0.6 (0.5)
Benefits would be reduced but not to $0 (consistent with offset rules)	**48.0**	**51.9**	36.7	11.3*** (1.7)	15.2*** (1.8)	3.9** (1.5)
Benefits would be reduced to $0 (consistent with current law)	43.2	39.8	**53.9**	-10.6*** (1.8)	-14.1*** (1.8)	-3.5** (1.5)
Benefits would neither stay same, nor be reduced	1.6	0.9	1.3	0.3 (0.3)	-0.3 (0.3)	-0.6* (0.3)
Don't know whether benefits would change	5.4	5.0	6.3	-0.8 (0.7)	-1.3 (0.7)	-0.4 (0.7)

Source: BOND Stage 2 36-Month Survey.

Note: The correct answers expected from T21, T22 and C2 subjects are indicated in bold. For example, "benefits would be reduced but not to $0" is the correct answer expected from T21 and T22 subjects (Appendix Section A.4.1). Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national population of SSDI-only beneficiaries who would volunteer for study enrollment. Standard errors are in parentheses.

Unweighted sample sizes: T21 = 3,785; T22 = 2,384; C1 = 3,661.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test with 9 degrees of freedom (resulting from a research design involving 10 study sites) and with no multiple- comparisons adjustment.

3.5.   Summary

The standard counseling offered to the Stage 1 treatment group and one Stage 2 treatment group was comparable to WIPA services offered to control subjects and non-BOND beneficiaries, as designed. Stage 1 treatment subjects' receipt of counseling services was slightly higher than for non-BOND beneficiaries, as would be expected because of several demonstration-related circumstances. Compared to standard counseling, the enhanced counseling offered to the Stage 2 T22 treatment group achieved the intensity benchmarks specified in the design of BOND. Enhanced counseling had statistically significant, positive effects on the use of employment services provided by state VR agencies and ENs.

Current law rules and offset rules about how earnings affect benefits are complicated. Despite the availability of counseling to all beneficiaries, understanding of how earnings affect benefits is low under current law (54 percent for the nationally representative Stage 1 control group). Efforts to inform Stage 1 treatment subjects led to 29 percent of the Stage 1 treatment group demonstrating correct knowledge of how earnings affect benefits under the offset rules three years after random assignment. The relatively limited understanding among the Stage 1 treatment group of how earnings affect benefits under the offset rules likely serves to constrain the behavioral response to the offset at least somewhat (although to what extent is unknown).

The evidence indicates that, relative to Stage 1 treatment subjects, more Stage 2 treatment subjects understood the offset. About half of the Stage 2 treatment subjects had a basic understanding of the offset. The Stage 2 control subjects had no better understanding of current law than the Stage 1 control subjects. Compared to the standard level of counseling, enhanced counseling resulted in a statistically significant, but not large increase in the percentage with a basic understanding of the offset rules (52 percentage points compared to 48 percentages points, i.e. an 8 percent increase).

4.   Using the Benefit Offset

To use the offset, treatment subjects must earn above the BOND Yearly Amount (BYA) after completing the trial work period (TWP) and grace period (GP). At that point, subjects receive a reduced benefit according to the offset rules. Timely and accurate benefit adjustments require that SSA and the implementation team complete multiple, complex processes. This chapter reports on the extent to which BOND treatment subjects used the offset and describes the characteristics of offset users. It also analyzes the processes that SSA and the implementation team used to make benefit adjustments.

Specifically, Section 4.1 shows what fraction of Stage 1 and Stage 2 treatment subjects used the offset through December 2016. Section 4.2 describes how offset use increased during the follow-up period and Section 4.3 describes patterns of offset use. Section 4.4 describes characteristics of offset users and Section 4.5 summarizes problems the demonstration encountered in implementing the offset rules to the treatment subjects and presents statistics on delays to benefit adjustments. Finally, Section 4.6 presents statistics on overpayments made to treatment subjects, which are a consequence of delays in the adjustment of benefits.

Exhibit 4-1 compares SSDI earnings rules under current law with rules under the benefit offset.

Exhibit 4-1.   Comparison of SSDI Current Law Rules to BOND Offset Rules

Both SSDI Current Law Rules and BOND Offset Rules
• Whenever SSDI beneficiaries work, they are required to report earnings to SSA. SSA also obtains evidence of earnings from the IRS and other sources.
• During the *Trial Work Period (TWP)*, earnings do not affect benefits. In 2017, a TWP month was any month in which an SSDI beneficiary had earnings of at least $840 or worked at least 80 self-employed hours. The TWP consists of nine such months in a rolling 60-month window.
• Given evidence of earnings, SSA conducts a *Work Continuing Disability Review (Work CDR)* to confirm beneficiaries' continued eligibility for benefit receipt. In SSA's terminology, disability "ceases" for beneficiaries who engage in substantial gainful activity (SGA) after completing the TWP.
• During the *Grace Period (GP)*, which starts with the disability cessation month and continues for two additional months of SGA, SSA pays benefits at their full amount regardless of earnings.

SSDI Current Law Rules	BOND Benefit Offset Rules
• After control subjects complete the TWP and GP, SSA suspends or terminates SSDI benefits in any month in which a beneficiary engages in SGA. Benefits are suspended for engagement in SGA in the first 36-months after the TWP, which is known as the re-entitlement period of the Extended Period of Eligibility (EPE). After the re-entitlement period, beneficiary engagement in SGA results in termination of entitlement to SSDI and the EPE ends; otherwise the EPE continues.	• After treatment subjects complete the TWP, they enter the BOND Participation Period (BPP), which continues for 60 months. During the BPP, after the beneficiary has used his GP, the $1 for $2 benefit offset applies to calendar-year earnings above the BOND Yearly Amount (BYA). After the BPP ends, beneficiary engagement in SGA results in termination of entitlement to SSDI.

4.1.   Few Stage 1 and Stage 2 Treatment Groups Used the Offset

A minority of beneficiaries in each treatment group used the offset. Through December 2016, 3.7 percent of Stage 1 treatment subjects used the offset (Exhibit 4-2). Not surprisingly, offset use is higher among

the Stage 2 treatment subjects, who represent all SSDI-only beneficiaries who would have volunteered had they been solicited for BOND (see Section 1.3). Specifically, 15.8 percent of those offered standard counseling and 15.4 percent of those offered enhanced counseling used the offset in this period.[36] Because of retroactive adjustments made after we extracted the data for this report in August 2017, actual use of the offset during the analysis period was somewhat higher.[37] Offset use estimates through 2015, the end of period aligned with the impact estimates in Chapter 5, are only slightly smaller than those for the period ending in 2016 (Exhibit 4-2).

To gain a better understanding of why offset use is limited to a minority of beneficiaries, we consider statistics for an important milestone on the way to use of the offset: engagement in SGA after TWP completion. To use the offset, treatment subjects must first work enough to complete the TWP and earn above SGA (Exhibit 4-1). SSA documents this milestone in administrative records using a cessation date. A cessation date is a necessary precondition for using the offset, so the fraction with a cessation date must be larger than the fraction using the offset. That said, for all treatment groups, the fraction with a cessation date is much larger than the fraction using the offset (Exhibit 4-2).

There are several reasons why the proportion of beneficiaries who used the offset may be lower than the proportion with a cessation date. First, some subjects stopped engaging in SGA sometime after their cessation date. In cases where the cessation date preceded BOND enrollment, beneficiaries may have stopped performing SGA before BOND enrollment.[38] Second, some adjustments for the period ending in December 2016 are still pending. Process statistics indicate, however, that the number of pending adjustments is now small (Section 4.5.1). Others may have reached their cessation date so recently in the period (essentially in the last quarter of 2016) that their first benefit adjustment will be in 2017— after the observation period.[39] More information about subjects who made progress toward the offset but did not use it is included in Volume 2, Appendix E, Exhibits E-11 and E-13.

---

[36]   The weighted statistics differ somewhat from unweighted statistics presented in past reports, where we focused on evaluating the demonstration processes rather than making inferences about national implementation. Weighted and unweighted statistics differ by small amounts. For example, the weighted offset use statistic for T1 in Exhibit 4-2 indicates is 3.68 percent (before rounding), whereas the corresponding unweighted statistic (not in the table) is 3.76 percent. The corresponding unweighted statistics for T21 and T22 are 15.70 and 15.55, again quite close to the weighted values reported in the exhibit.

[37]   The data supporting these statistics are from August 2017. Data from early October show that between August and October 2017, SSA retroactively adjusted benefits for an additional 0.4 percent of T1 subjects and 0.8 percent of Stage 2 subjects who used the offset by the end of 2016. This increase is due in large part to the automated reconciliation for 2016 earnings which took place in August 2017.

[38]   We found that 3.7 percent of T1 subjects (weighted) reached their cessation dates before they enrolled in BOND. Of those subjects, 35.2 percent went on to use the offset. For Stage 2, 11.2 percent of T21 subjects and 10.8 percent of T22 subjects reached their cessation dates before they enrolled in BOND, and of these 49.4 percent and 49.5 percent, respectively, went on to use the offset.

[39]   Treatment subjects continued to reach cessation dates in 2016. This includes 0.2 percent of T1 subjects, 0.6 percent of T21 subjects, and 1.5 percent of T22 subjects. Subjects with recent cessation dates may use the offset after the end of our analysis period.

Relative to WIC services, we did not find evidence that EWIC services increased offset use among Stage 2 volunteers (Exhibit 4-3 and Exhibit E-3). There were no significant differences in the percentages of T21 and T22 subjects using the offset in the full period of data available (2011-2016) or the time period aligned with the impact analysis (2012-2015) (Exhibit 4-3).

Exhibit 4-2.   Percentage of Treatment Subjects with Documented Steps towards Benefit Adjustment (through December 2016)

Time Period	2011-2015	2016 Additions	Total through 2016
Cessation Date Recorded in BTS			
T1	6.8	0.3	7.1
T21	24.0	0.6	24.6
T22	23.8	1.6	25.4
At Least One Year or a Partial Year of Offset Use			
T1	3.6	0.1	3.7
T21	15.2	0.6	15.8
T22	14.9	0.5	15.4

Source: BTS records from August 2017.

Unweighted Sample size: T1= 77,097; T21= 4,854; T22= 3,041

Note: We imputed cessation date values to account for rare anomalies in the BTS data. Specifically, less than one percent of Stage 1 treatment subjects successfully submitted an AEE and/or had an adjustment of benefits under the offset rules but did not have a cessation date recorded in BTS. The same is true for about one percent of all Stage 2 treatment subjects. Because a cessation date is a necessary step for successful AEE submission and benefit adjustment, we reclassified these beneficiaries as having a cessation date. These imputed cessation dates are included in the statistic presented in this exhibit. They represent 3.5 percent of all Stage 1 cessation dates, 4.2 percent of all T21 cessation dates, and 5.2 percent of all T22 cessation dates. Weights for Stage 1 subjects are used to ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment. Weights for Stage 2 subjects are used to ensure that the BOND subjects who met analysis criteria are representative the national population of SSDI-only beneficiaries who would volunteer for study enrollment.

Exhibit 4-3.   EWIC versus WIC Impact Estimates on Offset Use for Stage 2 Treatment Subjects

Outcome	Average Outcome with Offset and EWIC (T22)	Average Outcome with Offset and WIC (T21)	Estimated Impact of EWIC Instead of WIC Given Offset (T22 vs. T21)
Percentage of Beneficiaries With At Least One Year or a Partial Year of Offset Use			
Benefit offset use in 2011-2016	16.0	15.7	0.3 (1.0)
Benefit offset use in 2012-2015	14.4	14.4	-0.0 (1.0)

Source: SSA administrative records from the BTS (extracted August 2017) and MBR and the Stage 2 Baseline Survey.

Notes: We present two sets of estimates in this exhibit. The first, benefit offset use in 2011-2016, makes use of all the data available for this report. These estimates are based on SSA administrative records flagging any offset use through December 2016. We also present estimates of benefit offset use in 2012-2015, to align with the analysis period for the Stage 2 impact analysis presented in Chapter 5. These estimates are based on a different source: administrative records of the amount of benefit offset reduction made in each calendar year. These latter statistics intentionally exclude those Stage 2 treatment subjects who used the offset in 2011 but not in 2012 through 2015. They also, however, exclude any Stage 2 offset users with missing offset amounts in 2012 through 2015. Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national population of SSDI-only beneficiaries who would volunteer for study enrollment. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: T22 = 3,041; T21 = 4,854.

4.2.   Offset Use Increased in Each Year of BOND

Offset use increased in each year of the demonstration (Exhibit 4-4). The proportion of T1 subjects using the offset grew from 1.1 percent in 2011 to 2.2 percent in 2015 (Exhibit E-2). In Stage 2, the offset users grew from 6.1 percent of treatment subjects in 2012 (the first year when all subjects had enrolled in the study) to 9.7 percent in 2015 (Exhibit E-4).

Among those who used the offset in at least one year, the percentage using the offset increased in each year through 2015 (Exhibit 4-4). For example, of Stage 1 subjects who ever used the offset, 30.1 percent used it in 2011, 49.9 percent used it in 2013 and 62.3 percent used it in 2015. Growth in offset use (the change in the percent using the offset from one year to the next in Exhibit 4-4) was higher early in the demonstration and later tapered off.

Exhibit 4-4.   Offset Participation by Year among Treatment Subjects who Ever Used the Offset, Through December 2016
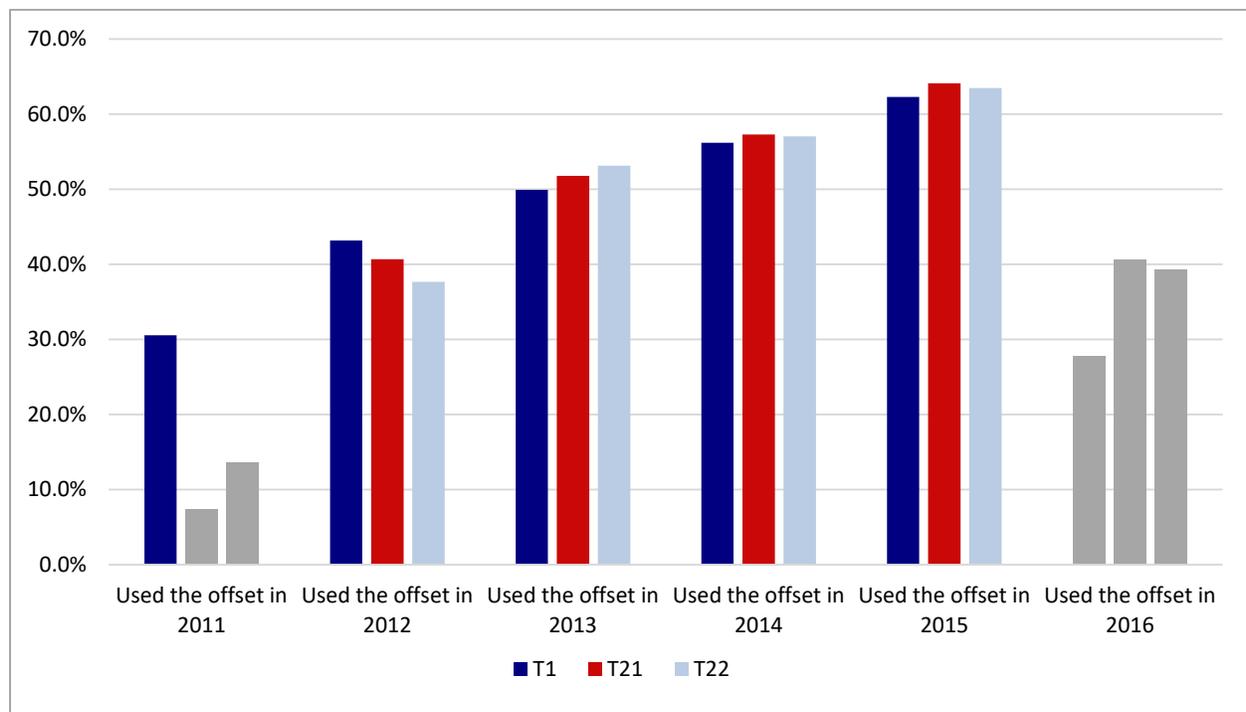
Source: BTS records.

Notes: The percentage shown for each year and treatment group is the percentage of all users from the 2011 through 2016 period whose benefits were reduced under the offset in the calendar year. See the right column in the bottom panel of Exhibit 4-2 for the percentage of each treatment group that used the offset in at least one month from 2011 to 2016. For 2011, the bars for Stage 2 subjects are gray because many Stage 2 subjects did not enroll in the demonstration until 2012; treatment subjects could not use the offset before enrollment. The 2016 percentages for all treatment groups are gray because they are likely to increase substantially after automated reconciliation of 2016 earnings, which had not yet occurred when we extracted data for this report. Use in earlier years may also increase because of retroactive adjustments, but by less substantial amounts than for 2016. Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the corresponding national beneficiary population. BTS information on years of offset use are missing for 59 T1 offset users, 28 T21 offset users, and 12 T22 offset users.

Unweighted sample sizes: T1 offset users: 2,836; T21 offset users; 734; T22 offset users: 461

There are two potentially important explanations for the observed changes in offset use over time. First, the offset rules may have induced some treatment subjects to seek substantial employment, and

presumably some of those subjects would take time to complete various steps before entering the offset. For treatment subjects in both stages, these steps potentially include some or all of obtaining training, finding work, completing the TWP and GP, and subsequently earning more than BYA. In addition, some Stage 1 subjects induced to earn enough to use the offset may only have learned about the opportunity late in 2014, as a result of the Stage 1 follow-up outreach effort. As a result, compared to offset use in the early years of BOND, offset use in later years may be more indicative of offset use in an ongoing national program, after offset rules have been in place for an extended period of time. Second, improving economic conditions (Exhibit C-2) likely increased beneficiaries' opportunities to work above BYA as the demonstration progressed.

What appears to be growth in offset use from 2011 to 2012 for Stage 2 treatment subjects is somewhat misleading; it is substantially due to the Stage 2 enrollment period, which extended from March 2011 through September 2012. Treatment subjects were not eligible to use the offset until they enrolled in the demonstration. All Stage 1 subjects were enrolled in May 2011, whereas the number of enrolled Stage 2 subjects increased throughout the 19-month enrollment period and only 40 percent had been enrolled by the end of 2011.

4.3.   A Few Treatment Subjects Used the Offset Early and Continuously, and a Majority of Offset Users Used it for Two or More Years

To better understand the extent to which beneficiaries were poised to use the offset when they enrolled in BOND, we examined how quickly offset users took advantage of the offset and how long they continued to use it. According to calculations based on BTS records from June 2017, 1.9 percent of Stage 1 subjects and 6.9 percent of Stage 2 subjects used the offset in 2011 or 2012. A substantial share of the subjects who began using the offset in 2011 or 2012 used the offset continuously through 2015 once they began (39.9 percent for those in Stage 1 and 36.0 percent of those in Stage 2).[40] However, those who first used the offset in 2011 or 2012 and used it continuously thereafter constitute a minority of all offset users through 2015 (25.0 percent of Stage 1 offset users and 19.1 percent of Stage 2 offset users).

Expanding the analysis of duration to include all those who started using the offset in 2013 or 2014, we find evidence that many of those who used the offset did so for two or three consecutive years.[41] Among all subjects who first used the offset between 2011 and 2014, a substantial majority (65.3 percent for Stage 1 and 62.0 percent for Stage 2) continued to use the offset for at least two consecutive years, and nearly half of offset users (47.2 percent of Stage 1 and 46.6 percent of Stage 2) used it for at least three consecutive years (Exhibits E-9 and E-10). Presumably some of these subjects continued to use the offset for four or more consecutive years, but the observation window for many is too short for us to assess.[42] In

---

[40]   We end the analysis of continued offset use in 2015 because it is the most recent year with comprehensive data on offset use. We expect offset use in 2016 to increase after automated reconciliation of 2016 earnings, which had not taken place at the time data for this report were pulled.

[41]   In the previous analysis our goal was to measure continuous offset use using the most complete years of data available (2011-2015). In this analysis our goal was to maximize the period of observation, despite the fact that data on offset use in 2016 will increase due to retroactive adjustments after SSA runs automated reconciliation for earnings in that year.

[42]   To illustrate this constraint, consider a beneficiary who first used the offset in 2014. In order to assess whether that beneficiary had used the offset for four consecutive years, we would need to know offset status in 2014

addition to continuous offset use, sizeable minorities of each treatment group stopped using the offset but returned to using it in a later year: 14.4 percent of Stage 1 offset users, and 16.8 percent of Stage 2 offset users used the offset for two or more non-continuous calendar years.

At face value, multiple years of offset use are indicative of both willingness and ability to use the offset for a long period. However, there are several reasons to believe that some beneficiaries initially make this choice without full understanding of the offset or awareness of their offset use. Approximately two-thirds of Stage 1 treatment subjects and one half of Stage 2 treatment subjects lacked an accurate understanding of the offset rules (Section 3.4). Due to backlogs in SSA processes, a substantial number had been using the offset for two or more calendar years before receiving a retroactive adjustment (Section 4.5.2), and they may not have fully recognized how the offset affected their income. The fact that they used the offset in multiple calendar years does not necessarily mean that their multiple-year use was the result of well-informed decisions.

4.4.   Several Beneficiary Characteristics Predict Offset Use

Among Stage 1 subjects, younger age and status as an SSDI-only beneficiary (rather than being a concurrent SSDI and SSI beneficiary) are statistically significant predictors of benefit offset adjustment (Exhibit E-12). Holding other characteristics constant, beneficiaries ages 20–29 were 10.4 percentage points more likely to use the offset relative to beneficiaries ages 55 and older, and SSDI-only beneficiaries are 1.5 percentage points more likely than concurrent beneficiaries to do so. Disabled adult child beneficiaries and beneficiaries with representative payees are less likely to have used the offset than beneficiaries without those characteristics, holding other characteristics constant.[43] In addition, several primary impairments were associated with a lower likelihood of offset use relative to those in the "other impairments" category: back or other musculoskeletal, nervous system disorders, circulatory system disorders, respiratory, and severe visual impairments.

Among Stage 2 subjects, younger age is predictive of offset use, and disabled adult child beneficiaries and beneficiaries with representative payees are less likely to have a benefit adjustment, holding other characteristics constant (Exhibit E-14), qualitatively the same as for Stage 1. In addition, Stage 2 subjects who, at baseline, had received SSDI benefits for 36 months or less at BOND entry ("short-duration" beneficiaries), had higher AIME, were working for pay or looking for work, were not in fair or poor health, or had a bachelor's degree or higher were more likely to have used the offset than others, holding other characteristics constant. Those with a respiratory primary impairment were more likely to use the offset than those in the "other impairments" category, while those with severe visual impairments were less likely to use the offset. Differences between the findings for Stage 1 and Stage 2 may be due to both

---

through 2017. However, we drafted the report in 2017 and final offset status in the current year is not determined until after the year is complete.

[43]   For each stage, we estimated a linear probability model for any offset use from through December 2016. The model for both stages includes explanatory variables for gender, age category, primary impairment, duration of SSDI receipt, and benefit amount and status, all at enrollment and based on administrative data. For Stage 2 subjects, the model also includes local economic conditions and explanatory variables from the Stage 2 baseline survey that were not available for Stage 1 subjects: employment status, educational attainment, and health status.

the differences in the populations they represent and the inclusion of survey-based baseline characteristics in the Stage 2 model.

4.5. Key Implementation Challenges Delayed Initial Benefit Adjustments

Problems with the implementation of the benefit adjustment process for offset users contributed to substantial delays of the first benefit adjustment. Instead, many offset users continued to receive full benefits when they should have received partial benefits. Thus, the delays in initial benefit adjustments may have reduced beneficiary understanding of offset rules and commonly led to overpayments.

BOND's design includes the following milestones on the administrative pathway to the first benefit adjustment:

1. *Sustained earnings sufficient for offset use:* To receive a benefit adjustment under the offset, treatment subjects must have sufficient sustained earnings to complete the TWP and GP followed by calendar-year earnings that exceed BYA. We refer to any demonstration year in which countable earnings exceed BYA after TWP and GP completion as a year with "offset use," even though the actual adjustment of the monthly benefit may not occur until late in the year or retroactively after the year has ended.

2. *Work CDR completion:* SSA must complete a work CDR to evaluate a beneficiary's work history, determine whether or not the beneficiary has completed the TWP and subsequently engaged in SGA, and establish when this occurred. There are three steps in the work CDR process: (1) SSA or BOND staff identify those in need of a work CDR based on beneficiary-reported earnings or information from other sources, typically an SSA-initiated review of IRS records; (2) beneficiaries, often with the help of SSA or BOND staff, have the opportunity to compile information on their work histories; and (3) SSA verifies the information and completes the work CDR.

3. *AEE submission:* Treatment subjects must provide an annual earnings estimate (AEE), an estimate of anticipated earnings during the calendar year. The BOND implementation team submits the AEE to SSA. SSA enters the information into BSAS to generate benefit adjustments.

4. *First benefit adjustment:* SSA's BOND Stand Alone System (BSAS) uses information on earnings in the calendar year to adjust SSDI benefits according to the benefit offset rules. SSA usually makes the first benefit adjustment later than, and retroactive to, the start of the year (or partial year) of offset use. When benefit adjustments are made retroactively, it typically means there has been an overpayment of benefits during the prior period of offset use. Delays in adjustment may cause underpayments for treatment subjects who were in the EPE and had their benefits suspended before BOND random assignment.

Offset users follow one of two different paths to the first benefit adjustment. We describe subjects who comply with the SSA requirement to report their earnings to the demonstration as "front door" entrants into the offset. For others, SSA discovers unreported earnings by a different path. We call these subjects

"back door" entrants into the offset. We show examples of each pathway in Exhibit 4-5.[44] Back-door entry into the offset requires more time for SSA to identify offset use and implement the first benefit adjustment (see Section 4.5.2).

Exhibit 4-5   Front-door versus Back-door Entry into the Offset

Step Toward First Benefit Offset Adjustment	Front Door	Back Door
Sustained earnings sufficient for offset use	Beneficiary proactively reports earnings to the demonstration	SSA discovers beneficiary's earnings through periodic review of IRS earnings data
Work CDR completion	SSA completes the work CDR, using information contributed by the beneficiary and assigns a disability cessation date	SSA completes the work CDR using information from non-beneficiary sources (e.g. the employer) and assigns a disability cessation date
AEE submission	Beneficiary provides an AEE before initial benefit adjustment	Beneficiary does not provide an AEE before initial benefit adjustment
First benefit adjustment	SSA uses information on calendar year earnings from the AEE to adjust benefit for the remainder of the calendar year, and to retroactively adjust benefits paid earlier in the calendar year	SSA adjusts benefits retroactively for a previous calendar year, after obtaining earnings information for that calendar year and completing a reconciliation of earnings and benefits

Note: By definition, the 'first benefit adjustment' occurs in the month when SSA first adjusts benefits, not in the month of first offset use. In most instances, the first benefit adjustment occurs after the month of first offset use.

4.5.1.   Key Implementation Challenges

Challenges to completing the administrative milestones have contributed to delays in identifying offset users and implementing the first adjustment of benefits under offset rules (see Exhibits E-5 through E-7). Timely first benefit adjustments were hampered in three notable ways: delays in completing work CDRs; delays in running automated end-of-year reconciliations of earnings and benefits; and early challenges with the submission of timely AEEs.

Delays in Work CDR Completion
The start of the work CDR process was delayed for beneficiaries who did not report earnings to the demonstration or to SSA despite SSA's requirement to do so; i.e., entered the offset through the "back door." The work CDR process begins when SSA or BOND staff identify beneficiaries in need of a work CDR based on beneficiary earnings. Beneficiaries who report earnings proactively can be identified as

---

[44]   In these examples we assume that beneficiaries' compliance or non-compliance with earnings reporting carries through to later steps in the benefit adjustment process. In practice, a front-door beneficiary (first column of Exhibit 4-5) could begin by reporting earnings but later follow steps we associate with back-door entry (second column), and vice versa.

needing a work CDR sooner than others. SSA discovers earnings for non-reporting beneficiaries after the fact in SSA-initiated reviews of IRS records.

The completion of the work CDR process was delayed by an insufficient number of the staff assigned to conduct this specialized work in SSA's ORDES BOND work unit. As a result, a backlog of pending work CDRs for treatment subjects accumulated over the first five years of the demonstration. To some extent, BOND inherited work CDR delays because some BOND subjects were already overdue for work CDR evaluation before the demonstration began.[45] By December 2015, 71 percent of BOND treatment group work CDR cases were over 270 days old, according to statistics from SSA's eWork system. After ORDES gained additional staff dedicated to work CDR processing in spring 2016 this figure fell to 12 percent in December 2016, and to 5 percent in July 2017. Information from eWork suggests that there was no comparable work CDR backlog for control subjects. That source indicates that approximately one percent of beneficiaries subject to current law (including both control subjects and beneficiaries not in BOND) had work CDRs more than 270 days old at each of these points in time.

Delays in Automated End-of-Year Reconciliations
In addition, ongoing problems with BSAS functionality diverted staff time from work CDR processing and delayed first adjustments for back-door offset users.[46] SSA delayed automated end-of-year reconciliations of earnings and benefits for 2011, 2012 and 2013 due to problems with BSAS functionality (Derr et al. 2015).[47] As a result, BSAS was unable to automatically process a significant number of end-of-year reconciliations. Instead, ORDES staff had to conduct manual reconciliations; that is, SSA staff (rather than BSAS) processed the end-of-year reconciliations. Even after staff ran automated reconciliation on time for 2014 earnings and benefits, the problems requiring manual end-of-year processing remained. This continued to divert ORDES resources from work CDR processing (Croake et al. 2017).[48] The late runs of BSAS also meant that some beneficiaries who did not submit AEEs experienced an additional delay to their first benefit adjustment. This happened when SSA identified offset use in a year for which it had not yet run automated reconciliation. If the beneficiary did not submit

---

[45]   In fiscal year 2010—before BOND began enrolling subjects—SSA took 124 days to process work CDRs on average (SSA 2011). Between January and April 2011, SSA made an effort to complete pending work CDRs for future Stage 1 subjects ahead of their random assignment into BOND. However, it was not possible to complete all of the pending work CDRs during this time.

[46]   In addition to using BSAS to adjust benefits following the submission of an AEE, after the end of each calendar year, SSA also uses BSAS for an automated reconciliation process that compares estimated earnings to earnings reported in IRS records and makes additional retroactive benefit adjustments for the prior year in the event of a substantial difference.

[47]   Automated reconciliation for a given year is scheduled for August of the following year. SSA delayed the 2011 automated reconciliation by five or six months (conducted in January and February 2013) and the 2012 automated reconciliation by one or two months (conducted in September and October 2013). SSA conducted the 2013 automated reconciliation in late April through May 2015—an eight month delay.

[48]   SSA uses automated reconciliation to review earnings and benefits for all treatment subjects with cessation dates. During the most recent automated reconciliation (run in August 2016 for 2015 earnings), BSAS could not fully process the vast majority of beneficiaries' cases (about 3,000 of the approximately 3,700 cases; 78 percent). The same was true for the majority of beneficiaries whose cases were processed during the 2015 automated reconciliation. These cases required at least some level of manual processing by ORDES staff.

an AEE, SSA waited to adjust benefits until completion of automated reconciliation for the year. Finally, delays in reconciling earnings and benefits for 2011, 2012 and 2013 led to an extended period of time in which subjects may have unknowingly accumulated over- and underpayments (Section 4.6).

Delays in AEE Completion

Early in the demonstration, problems with the timely completion of AEEs by some counselors also contributed to delays in first benefit adjustments. This process has run smoothly since 2013, after the implementation team began reviewing BTS data to identify beneficiaries in need of an initial AEE, and shifted the majority of post-entitlement work from benefits counselors to a centralized team (Derr et al. 2015). This process was not uniform across treatment arms. In fact, we found evidence consistent with the hypothesis that EWIC services accelerated beneficiaries' submission of AEEs. By 2014, the percentage of T22 subjects with submitted AEEs was a statistically significant 2.8 percentage points higher than for the T21 subjects, but by the end of 2016 that difference had diminished to an insignificant 1.2 percentage points (Exhibit E-3).

4.5.2.   Median Duration From First Offset Use to First Benefit Adjustment

To understand the aggregate effect of factors affecting the speed of initial adjustments, we examine the duration from first month of offset use to the month in which SSA makes the first adjustment. Ideally, SSA would first adjust benefits in the first month of offset use, or shortly thereafter. The first month of offset use is often the third month after the disability cessation month, coinciding with GP completion. Thus, after the cessation month occurs, SSA has a two-month window to identify the cessation month and make a timely adjustment.

Across all groups of treatment subjects, many first benefit adjustments were implemented well beyond the first month of offset use (Exhibit 4-6). The median duration from offset use to first benefit adjustment ranged from 13 to 22 months, depending on the treatment group.[49]

As described in Section 4.5.1, delays in benefit adjustment can stem from beneficiary failure to report earnings to the demonstration or to SSA. We expected that beneficiaries who proactively report their earnings and otherwise engage with the demonstration would have a shorter duration from first offset use to first benefit adjustment. Such beneficiaries provide a best-case-scenario for the adjustment process. To examine this, we measured median processing times (Exhibit 4-6) for several partially overlapping groups of beneficiaries, based on indicators of their engagement with the demonstration to facilitate timely adjustment. The first group represents the most proactive treatment subjects: those who reported their earnings to demonstration staff before their first month of offset use. (This group does not include those who reported earnings directly to SSA, as we do not have information on direct reports to SSA.) The second group includes those who entered the offset via submission of an AEE. In contrast to the balance of beneficiaries who entered the offset via SSA-initiated reconciliation,[50] submission of an AEE reflects

---

[49]   Unlike other statistics in this chapter, we do not use analytic weights to adjust statistics related to wait times for benefit adjustment. Rather, we present the actual experiences of treatment subjects under the demonstration.

[50]   SSA-initiated reconciliation includes automated end-of-year reconciliation and manual reconciliation. SSA initiates manual reconciliations for calendar years in which SSA already completed automated reconciliation. See Section 5.2.2 of Derr et al. (2015) for more details.

some engagement with the demonstration before benefit adjustment. We also examine processing times for the union of these two groups and for beneficiaries who are in neither of these two groups.

Consistent with the expectations, we find that engagement with the demonstration to report earnings or submit an AEE is associated with a shorter duration between first offset use and first benefit adjustment. We find that the median duration to initial adjustment was shorter for offset users who reported earnings or had an initial adjustment through an AEE than for those who did not conduct either activity, by 5 months for T1, 6 months for T21, and 11 months for T22. The shortest adjustment times were observed among those who reported earnings to BOND before first offset use. Even so, subjects who reported their earnings to the demonstration experienced delays to the first benefit adjustment, with median durations to initial adjustment of 12 months for T1, 14 months for T21 and 8 months for T22.

The fact that, among all offset users, the median durations for both Stage 2 groups are shorter than for Stage 1 likely reflects the fact that Stage 2 groups are volunteers who completed an informed consent process which may have increased their awareness of the beneficiary's role in timely benefit adjustment. Indeed, comparing the Stage 1 and Stage 2 treatment groups with access to standard counseling, we see that 52 percent of T1 offset users reported earnings to BOND before first offset use or had an initial adjustment through an AEE compared to 85 percent of T21 offset users. Overall, benefits were adjusted 5 months faster for the T21 offset users.

We found the shortest duration to first benefit adjustment in the T22 group. Assignment to T22 accelerated adjustments relative to assignment to T21. EWIC outreach may have accelerated recognition of the need to report and submit an AEE to SSA. Indeed, 90 percent of T22 subjects reported earnings before first offset use or had an initial adjustment through an AEE. Among those who engaged with the demonstration to facilitate adjustment, adjustment times were also shorter for T22 subjects relative to T21 subjects. The T22 subjects may have reported their earnings earlier or completed an AEE faster relative to T21 subjects, potentially with the guidance or urging of EWIC staff.

Exhibit 4-6.   Median Months from First Offset Use to First Benefit Adjustment

	T1	T21	T22
All Offset Users Through 2016	22.2	17.3	13.2
Reported earnings to BOND before first offset use or initial adjustment through an AEE	17.9	15.3	10.3
Reported earnings to BOND before first offset use	12.3	14.1	8.4
Initial adjustment through an AEE	17.4	14.2	8.9
Did not report earnings to BOND before first offset use and initial adjustment through SSA reconciliation	22.9	20.9	21.6

Source: BTS records.

Note: BTS started to track the date of initial benefit offset adjustment on February 25, 2013. Hence, we do not have necessary information to calculate the duration from first offset use to first benefit adjustment for the 11 percent of Stage 1 adjustments and 15 percent of Stage 2 adjustments that occurred before February 25, 2013.

Unweighted sample sizes: All T1 offset users: 2,610; T1 subjects who reported earnings to BOND before first offset use or initial adjustment through AEE: 1,352; T1 subjects who did not report earnings before first offset use and with initial adjustment through SSA reconciliation: 1,258; all T21 offset users: 660; T21 subjects who reported earnings to BOND before first offset use or initial adjustment through AEE: 560; T22 subjects who did not report earnings before first offset use and with initial adjustment through SSA reconciliation: 100; all T22 offset users: 404; T22 subjects who reported earnings to BOND before first offset use or initial adjustment through AEE: 362; T22 subjects who did not report earnings before first offset use and with initial adjustment through SSA reconciliation: 42

Delays in first benefit adjustments may reduce beneficiary understanding of offset rules. Many offset users did not fully experience the income consequences of their earnings until they received the first benefit adjustment 1 to 2 years into the five-year BPP. For these beneficiaries, the nature of the connection between earnings and benefit adjustments may have been obscured by the long delay. We do not know whether delays reduce beneficiary understanding of the offset. We have been able to assess whether beneficiary understanding of the offset improves when SSA makes an adjustment, but have not found that it does.[51]

Another implication of delayed first adjustments is that they increase the incidence of overpayments. This is because offset users continued to receive full benefits when they should have received partial benefits. We discuss overpayments in the next section (Section 4.6).

4.6.   Overpayments

Delays in timely and accurate benefit adjustment may result in SSA issuing improper payments to beneficiaries. In BOND, offset use was often accompanied by overpayments. In this section, we define

---

[51]   To investigate how beneficiary understanding of the offset changes after the first adjustment, we studied a small sample of Stage 2 offset users who responded to the 12-month and 36-month surveys and received their first adjustment between the two surveys. We compared their responses on a question designed to assess understanding of the benefit offset earnings rules, and found that understanding was no better after the initial adjustment (Exhibit E8). However, we do not know if more timely benefit adjustment would have improved understanding of the offset rules.

overpayments, describe treatment beneficiaries' experiences with overpayments, and present estimates of the benefit offset's impact on overpayments.

4.6.1.   Definition of Overpayments

Work-related overpayments occur when SSA pays beneficiaries more in SSDI benefits than the amount to which they are entitled on the basis of work activity. Under current law, beneficiaries may be overpaid because of delays in beneficiary reporting of earnings and SSA processing of earnings information. For current-law beneficiaries who are overpaid because they engage in SGA, monthly overpayment amounts are typically equal to the entire (monthly) benefit check. This is because their benefits should be suspended or terminated—they have completed the TWP and GP and continued to engage in SGA.

BOND treatment subjects may also encounter work-related overpayments. Similar to control subjects, treatment subjects may be overpaid because of delays in the reporting and processing of earnings information. In these instances, overpayments will be no larger than, and often smaller than what their overpayment would have been under current law. This is because they may be entitled to partial benefits under the BOND offset rules.

Treatment subjects may also be overpaid for reasons that do not apply to control subjects. One reason is that their monthly benefits are based on annual, rather than monthly, earnings. Several components of administering an annual accounting period create the opportunity for an overpayment. This includes BSAS errors, inaccurate AEEs, and revised AEEs. Section 8.1 of Hoffman et al. (2017) provides more information on the causes of overpayments for treatment subjects and also their resolution.

Some overpayments occur for reasons other than work, but are not relevant to the evaluation of BOND.[52] That is, we do not expect BOND to affect the prevalence or size of overpayments that occur for reasons other than work. Hence, we focus on work-related overpayments. For brevity, we use the term overpayments to refer to work-related overpayments.

4.6.2.   The Overpayment Experience of Offset Users

Offset use was accompanied by overpayments for the large majority of Stage 1 offset users. We produce overpayment estimates for the subsample of BOND disabled-worker beneficiaries whose records contained sufficient data for analysis (see Appendix E.1.1).[53] Among this disabled-worker sample, 89 percent of T1 offset users had an overpayment during the BOND analysis period (Exhibit E-15). The average total overpayment amount (which is inclusive of all overpayment accrued during the analysis period) among overpaid T1 subjects was $7,317 (Exhibit E-15). This is a sizable amount, approximately seven months' worth of an average SSDI benefit check.

---

[52]   Work-related overpayments accounted for 31 percent of total SSDI overpayment dollars between late 2003 and early 2014. These overpayments represented the second highest share of total overpayment amounts, behind overpayments related to medical improvement (SSA Office of Inspector General 2015).

[53]   The overpayment analysis sample includes about 82 percent of Stage 1 subjects and 92 percent of Stage 2 subjects. We exclude disabled adult child and disabled widow beneficiaries from the overpayment analysis. Such beneficiaries account for almost 15 percent of the Stage 1 sample and 5 percent of the Stage 2 sample and comprise the bulk of beneficiaries excluded from the overpayment analysis.

The prevalence and size of overpayments made to Stage 1 offset users differed before and after their first benefit adjustment. Most—83 percent of Stage 1 offset users—experienced an overpayment while using the offset in the period before SSA implemented their first benefit adjustment (Exhibit E-16). Thus, overpayments stemming from delays in initial beneficiary reporting and SSA processing of earnings information were highly prevalent. After the first benefit adjustment, a smaller percentage (58 percent) of offset users experienced overpayments and mean overpayments were lower than those experienced before first adjustment (Exhibit E-16). At least in part, this is because once SSA completes a work CDR and identifies offset use, it suspends benefits until the beneficiary submits an AEE. Overpayments experienced after first adjustment may have occurred for reasons related to the annual accounting period such as incorrect AEEs.

Similar proportions of T21 and T22 subjects were overpaid, but overpayment amounts differed by treatment group. Between 2012 and 2015, 91 percent of T21 offset users and 92 percent of T22 offset users were overpaid (Exhibit E-17). The mean total overpayment amount was $6,068 among overpaid T21 subjects and $4,797 among overpaid T22 subjects (Exhibit E-17). The accumulation of overpayment debt reflects the duration from first offset use to first benefit adjustment presented in Exhibit 4-6. T22 subjects had the shortest duration and also experienced the lowest overpayment amounts. Nevertheless, overpayments were prevalent and sizable even among recruited and informed volunteers who received proactive counseling, presumably because of systemic issues in the BOND processes for gathering and processing earnings information.

When SSA identifies an overpayment, it requires beneficiaries to repay the owed amount either by check or through withheld future benefits. Not surprisingly, some, but not all, beneficiaries have negative reactions. WIC and EWIC counselors described beneficiaries' reactions to overpayments as generally negative. It is possible that beneficiaries with negative experiences with overpayments were more likely to express their reactions to WIC and EWIC counselors compared to beneficiaries with neutral reactions. Indeed, beneficiaries reported a wider range of reactions to overpayments during in-depth interviews in 2015. Consistent with the counselors' perceptions, some beneficiaries described negative reactions and adverse financial outcomes following overpayments. Several beneficiaries, however, described neutral responses to overpayments. Beneficiary reactions may vary based on the size of the overpayment, the beneficiary's financial status, awareness of a pending overpayment, and previous experience with overpayments, among other factors.

Of course, overpayments are not unique to BOND treatment subjects. Below, we present estimates of the impacts of the benefit offset on both the prevalence and mean size of overpayments during the demonstration period.

4.6.3.   Impacts on Overpayments

BOND's experimental design supports a rigorous analysis of the impact of the offset on the rate and size of overpayments. As with other impact estimates, we estimate impacts for all beneficiaries in the treatment group (excluding a small share due to data issues), not just those who used the offset.

In the previous subsection, we discussed the prevalence of overpayments and mean overpayments among the subset of offset users. The impact analysis focuses on the broader group of all BOND subjects, regardless of offset use. Because only a minority of BOND subjects uses the offset and, hence, only a minority are at risk of overpayment, the prevalence and mean size of overpayments among all BOND treatment subjects is lower than the statistics presented for offset users.

We expect that the offset will have a positive impact on the prevalence of overpayments for three reasons. First, SSA was delayed in processing work CDRs for treatment subjects relative to control subjects. Delays in work CDR completion postpone benefit adjustment and may lead to overpayments. Second, treatment beneficiaries may have encountered additional delays in benefit adjustment due to delays in automated reconciliation. Delays to automated reconciliations under the BOND offset did not delay benefits to control subjects; SSA reconciled their earnings three times per year. As a result, SSA may have retroactively adjusted benefits and notified beneficiaries of their status more timely for control subjects relative to treatment subjects. Third, under the offset, benefits are based on an estimate of annual earnings and require retroactive adjustment whenever actual earnings differ by more than a minimal amount.

As expected, we find that the offset increased the prevalence of overpayments among treatment beneficiaries. The prevalence of overpayments was higher among treatment subjects than among control subjects, by 25 percent of the control group mean for Stage 1 and by 54 percent for Stage 2 (calculations based on Exhibits 4-7 and 4-8). For Stage 2, there was no statistically significant difference in the impact of assignment to T22 rather than T21 on the prevalence of overpayments.

Exhibit 4-7.   Estimated Impacts on Overpayments among Stage 1 Subjects in 2011 to 2015

Outcome	T1 Mean	C1 Mean	Impact Estimate
Percentage of Stage 1 Subjects with Overpayment			
Overpaid in any month in 2011, 2012, 2013, 2014, or 2015 (%)	3.42	2.74	0.68*** (0.11)
Mean Overpayment Amount for All Stage 1 Subjects			
Combined 2011, 2012, 2013, 2014, and 2015 overpayment	$250	$348	-$98*** ($14)

Source: Monthly DBAD extracts from May 2011 to December 2015 and May 2017 and baseline SSA administrative data.

Note: We estimate overpayments starting in May 2011. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W). Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment. This analysis includes only BOND disabled-worker beneficiaries whose records contained sufficient data for analysis (see Appendix E.1.1). Unweighted sample sizes: T1 = 65,127, C1 = 716,403.

Exhibit 4-8.   Estimated Impacts on Overpayments among Stage 2 Subjects in 2012 to 2015

Outcome	T21 Mean	T22 Mean	T21+T22 Mean	C2 Mean	Estimated Impact of Offset vs. Current Law (T21+T22 vs. C2)	Estimated Impact of EWIC vs. WIC (T22 vs. T21)
Percentage of Stage 2 Subjects with Overpayment						
Overpaid in any month in 2012, 2013, 2014, or 2015 (%)	13.8	13.5	13.3	8.6	4.7*** (0.76)	-0.3 (0.88)
Mean Total Overpayment Amount for All Stage 2 Subjects						
Combined 2012, 2013, 2014, and 2015 overpayment	$837	$646	$753	$1,030	$-277** ($94)	$-191** ($81)

Source: Monthly DBAD extracts from January 2012 to December 2015 and May 2017 and baseline SSA administrative data.

Note: We estimate overpayments starting in January 2012. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W). Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national population of SSDI-only beneficiaries who would volunteer for study enrollment. This analysis includes only BOND disabled-worker beneficiaries whose records contained sufficient data for analysis (see Appendix E.1.1).

Unweighted sample sizes: T21 = 4,468, T22 = 2,787, C2 = 4,447.

Although treatment subjects were more likely than control subjects to have an overpayment, for both stages, mean overpayment amounts were significantly lower than for control subjects. Mean overpayment amounts were 27 to 28 percent lower for Stage 1 and Stage 2 treatment subjects relative to the amounts for corresponding control subjects (calculations based on Exhibits 4-7 and 4-8). The differences for the full Stage 1 and Stage 2 samples reflect a smaller overpayments among overpaid treatment beneficiaries than among overpaid control beneficiaries (Exhibits E-15 and E-17).[54]

Unlike with the prevalence of overpayments, we did find a significant difference in mean total overpayment amount across the T21 and T22 groups. The mean overpayment amount of the T22 group was $191 less than that of the T21 group (Exhibit 4-8). A plausible explanation is that outreach and assistance from EWIC staff helped some T22 subjects report earnings sooner and submit AEEs that were typically more accurate than those submitted by their T21 counterparts. Nonetheless, these efforts did not reduce the prevalence of overpayments to T22 subjects relative to T21 subjects.

4.7.   Summary

As of December 2016, 3.7 percent of T1 subjects and 15.6 percent of Stage 2 subjects had used the offset and received a benefit adjustment. Among Stage 2 subjects, the availability of EWIC services did not increase offset use compared to WIC services. For Stage 1 subjects, who are representative of all SSDI

---

[54]   We calculate the mean overpayment amount among overpaid beneficiaries as the mean overpayment among the larger group of beneficiaries (those with and without overpayments) divided by the prevalence of overpayments among the larger group of beneficiaries. For C1 subjects with overpayments, that calculation is $348.37/0.027415 = $12,707. The mean for overpaid T1 subjects is $7,317 (Exhibit E-15). For overpaid C2 subjects, that calculation is $1,030.41/.086122 = $11,965. The mean for overpaid Stage 2 treatment subjects is $5,662 (Exhibit E-17).

beneficiaries, younger age and SSDI-only status are predictors of offset use. Younger age is also predictive of offset use among Stage 2 subjects, in addition to several baseline predictors from survey data.

The number of Stage 1 and Stage 2 offset users grew each year of the demonstration. Some of this growth may simply be due to use by those who, at the outset of the demonstration, still had unused TWP and GP months. Some users may also have needed time to prepare for and obtain work that would allow them to use the offset. Some Stage 1 users may have learned about the offset several years after the demonstration started, from the follow-up outreach effort. For all users, the economic expansion over the period may have provided more opportunities to use the offset as the demonstration progressed. We expect that offset use in the later years of the demonstration is more indicative of what participation would be under a national program, assuming similar economic conditions—once all beneficiaries have had an opportunity to complete the steps on the path to offset use. In 2015, 2.2 percent of Stage 1 treatment subjects and 9.7 percent of Stage 2 subjects used the offset. Compared to earlier years, by 2015 more treatment subjects had both more time and more opportunities to adapt their behavior to the BOND offset.

Due to several implementation issues, most notably delays in SSA's work CDR processing and functionality of software supporting offset adjustments, for many offset users SSA first adjusted benefits one to two years after the month of first use. Median duration from first month of offset use to first benefit adjustment was 22 months for Stage 1 and 15 months for Stage 2. Some of the delays, especially for Stage 1, occurred because some subjects did not report their earnings to SSA. Although there are limitations on our ability to compare delays for those offset users who reported earnings and those who did not, available statistics indicate that the delays experienced by the former were typically several months shorter than for the latter, but still quite long. That is an important explanation for the longer median duration for Stage 1. For Stage 2, delays were typically shorter for T22 subjects than for T21 subjects, most likely because of the proactive outreach of EWIC staff. Delayed first adjustments may reduce beneficiary understanding of the offset and increase the incidence of overpayments.

In large part because of delays in benefit adjustments, overpayments were more prevalent among treatment subjects relative to control subjects. Offset use was coupled with overpayments for about 90 percent of offset users. Among all treatment subjects, the offset caused an increase in the prevalence of overpayments, but decreased mean overpayments accrued over the analysis period. The likelihood of an overpayment was 25 percent higher for Stage 1 treatment subjects than for C1 subjects, but the mean overpayment accrued was 28 percent less than for C1 subjects. The relative impact on prevalence for Stage 2 treatment subjects compared to C2 subjects is larger than the corresponding value in Stage 1, a 54 percent increase. However, the relative impact on mean overpayments is about the same as in Stage 1: a 27 percent reduction. Comparison of overpayments across the two Stage 2 treatment groups found no evidence that EWIC services affected the prevalence of overpayments relative to WIC services. However, we did find evidence that EWIC services reduced the mean total overpayment amount.

The estimated impacts of the offset on the relative prevalence and size of overpayments likely differ from what we would expect under a national benefit offset program. This is because work CDRs were delayed for treatment beneficiaries relative to beneficiaries subject to current law, particularly early in the demonstration. If treatment subjects experienced the same work CDR processing times as beneficiaries subject to current law, we would expect to see a lower prevalence and size of overpayments. This would result in smaller impacts of the offset on the prevalence of overpayment and larger (in absolute value) negative impacts on the size of overpayments.

5.   Impacts on Earnings, SSDI Benefits, and Other Outcomes

This chapter describes the impacts of the BOND benefit offset policy and enhanced counseling on a range of outcomes. The chapter begins with a description of the hypothesized impacts on the central earnings- and benefits-related outcomes and a summary of the main findings. It then turns to the detailed findings: estimated impacts on earnings-related and benefits-related outcomes measured from administrative data. The chapter also presents results for reporting of non-countable earnings, a key process outcome. The final section of the chapter reports impact findings for survey-derived outcomes.

5.1.   Expected Impacts and Summary of Main Findings

BOND tests whether replacing the SGA cash cliff with a $1 for $2 offset "ramp" would increase work and earnings and reduce the total amount of SSDI benefits paid to beneficiaries (Bell et al. 2011). In particular, the demonstration provides three tests of the benefit offset policy versus current law:

- Stage 1's T1 versus C1 comparison;
- Stage 2's T21 versus C2 comparison; and
- Stage 2's T22 versus C2 comparison.

Stage 2 of the demonstration provides the additional test of enhanced work incentives counseling (EWIC) versus standard work incentives counseling (WIC) for beneficiaries subject to the benefit offset rules:

- Stage 2's T22 versus T21 comparison.

For the benefit offset tests, the theoretical direction of impacts on mean earnings and mean benefits is ambiguous.[55] As discussed in Chapter 1, this ambiguity arises because the incentives created by the benefit offset vary with what a beneficiary's earnings would be under current law. Opposite effects are expected for those who, under current law, would have had earnings below BYA (hereafter the "below BYA group") and those who, under current law, would have had earnings above BYA (hereafter the "above BYA group").

- The offset is expected to induce treatment group subjects in the below BYA group to have on average (i) higher earnings and (ii) lower SSDI benefits than they would under current law. The lower SSDI benefits result from the change in earnings behavior for subjects induced by the offset to increase earnings above the BYA threshold.

- Conversely, the offset is expected to cause treatment subjects in the above BYA group to have on average (i) lower earnings (though still above BYA) and (ii) higher SSDI benefits than they would under current law. The higher SSDI benefits result from two causes. First, there is a

---

[55]   Total earnings and total SSDI benefits are the two predesignated confirmatory outcomes in the evaluation. These outcomes address the primary policy objectives for the benefit offset policy. We apply a higher standard of evidence to the confirmatory outcomes by adjusting the p-values of the impacts estimates to account for multiple comparisons. This adjustment ensures that the possibility of incorrectly rejecting "no effect" (the null hypothesis) on any outcome is no more than 10 percent within each stage.

mechanical effect, by which the offset rules provide partial benefits in months when current law would have suspended or terminated benefits. This provides most of those in the above BYA group with a windfall for no change in behavior.[56] Second, the increase in income from the partial benefits will induce some to lower their earnings. This change in earnings behavior serves to further increase SSDI benefits.[57]

The net impacts on earnings and benefits combine effects on these two types of subjects.

- For the impact on total earnings to be positive, the *positive impact* expected for the below BYA group would have to be *larger* than the *negative impact* expected for the above BYA group.

- Similarly, for the impact on SSDI benefits to be negative, the *reduction in benefits* for the below BYA group must be *larger* than the *increase in benefits* for the above BYA group.

If the benefit offset has no or little effect on earnings behavior, then the non-behavioral windfall effect will dominate, causing an increase in total SSDI benefits.

Unlike for earnings and SSDI benefits, theory does predict the signs of the impacts for several other outcomes—each of which we treat as exploratory (for discussion of confirmatory vs. exploratory outcomes see Chapter 2). Most notably, theory predicts positive impacts on employment (i.e., any positive earnings) and on the percentage with earnings above BYA. Other predictions for exploratory outcomes are described in Exhibit 5-1. This exhibit lists the central administrative outcomes and provides an explanation for the predicted sign of the impact (positive, negative, or ambiguous).

To preview the impact results, it is useful to divide the effects of the benefit offset policy into behavioral and non-behavioral effects. As noted above, if there is no change in behavior, earnings will be unchanged, but SSDI benefits will rise. The nationally representative Stage 1 results provide evidence of limited changes in earnings behavior in response to the benefit offset. While those behavioral responses appear consistent with the theory, they are small and in opposite directions. Thus, it would be reasonable to characterize the Stage 1 findings as "no substantively important impacts on earnings behavior".

Specifically, for Stage 1, we find no statistically significant impact on *mean earnings*. In addition, the point estimate of the impact is very close to zero. Relative to the control group, the benefit offset increased *SSDI benefits* over the full follow-up period of 56 months by about 1 percent of the control group mean. The magnitudes of the Stage 2 impact estimates imply larger behavioral responses for informed SSDI-only volunteers—those for whom we expected behavioral responses to be largest. Nevertheless, even with larger effects on employment and the share with earnings above BYA, the measured effects on average earnings in Stage 2 are not statistically significant. The positive impact on benefits found in Stage 2 appears to be largely driven by the same non-behavioral mechanism as in Stage 1. We also find no statistically significant incremental impacts of the counseling enhancement tested in Stage 2 for the earnings- and benefits-related outcomes.

---

[56] High-earning beneficiaries whose earnings above BYA are more than two times their benefit amount will be in "full offset" and so will not be due any SSDI benefits.

[57] The reduction in earnings by some above BYA group beneficiaries reduces the size of the benefit offset.

Exhibit 5-1.   Predicted Direction of Impacts on Earnings-Related and Disability Benefit-Related Outcomes

Outcome	Predicted Direction	Explanation of How Offset Policy will Affect Outcome
Confirmatory Outcomes		
• Total earnings in period[a]	?	Opposing effects for those who would earn below and above BYA under current law.
• Total SSDI benefits in period[b]	?	Opposing effects for those who would earn below and above BYA under current law.
Exploratory Outcomes		
Earnings-Related Outcomes[a]		
• Any employment in period (%) • Number of years with any employment	+	Would earn under BYA under current law: expected increase in average employment. Would earn above BYA under current law: no change in employment expected (expected to continue employment).
• Earnings above BYA in at least one year (%) • Number of years with earnings above BYA	+	Would earn under BYA under current law: expected increase in proportion earning above BYA. Would earn above BYA under current law: expected to continue to earn above BYA.
• Earnings above 2 times BYA in at least one year (%) • Number of years with earnings above 2 times BYA	?	Opposing effects— Would earn under BYA under current law: expected increase in average earnings. Would earn above BYA under current law: expected to reduce earnings because of income and substitution effects.
• Earnings above 3 times BYA in at least one year (%) • Number of years with earnings above 3 times BYA	?	Same explanation as 2 times BYA outcomes.
Disability Benefit Outcomes[b]		
• At least one month with SSDI benefits (%) • Number of months with SSDI benefits	+	Would earn under BYA under current law: no change (or slight reduction from moving beyond ramp) expected in months with SSDI benefits. Would earn above BYA under current law: move from suspended benefits to partial benefits increases months with SSDI benefits, unless earnings are so high that benefits are zero under the offset rules.
• Total SSI benefits in period • At least one month with SSI benefits (%) • Number of months with SSI benefits	–	For concurrent beneficiaries earning above BYA, partial SSDI benefits will reduce average SSI benefits. For a few, the partial SSDI benefits will lead to suspense of SSI benefits. For SSDI-only beneficiaries earning above BYA, higher income under offset policy should reduce the proportion that spends down assets enough to meet SSI resource test, reducing or slowing entry into SSI.

Notes: "+" = predicted positive impact. "–" = predicted negative impact. "?" = ambiguous predicted impact.

See Chapter 2 for discussion of confirmatory versus exploratory distinction. See the *BOND Evaluation Analysis Plan* (Bell et al., 2011) for an extended discussion of the theoretical predictions.

[a] For earnings-related outcomes, Stage 1 follow-up period is January 2011 to December 2015 and Stage 2 follow-up period is January 2012 to December 2015.

[b] For disability benefit outcomes, Stage 1 follow-up period is May 2011 to December 2015 and Stage 2 follow-up period is January 2012 to December 2015.

5.2.   Stage 1

This section first discusses the impact estimates on confirmatory outcomes. Then it discusses exploratory results: impact estimates on key exploratory outcomes, impacts for each year of the follow-up period, subgroup analysis, and impacts on non-countable earnings.

5.2.1.   Estimated Effects on Confirmatory Outcomes

The Stage 1 analysis finds no confirmatory evidence of an impact on total earnings over the five-year follow-up period (Exhibit 5-2). The statistically insignificant impact estimate on total earnings over five years is $9, less than 0.2 percent of the control group average ($6,622). The 95 percent confidence interval is narrow and ranges from -$197 to +$217, or from -3 to +3 percent of the control group average. Even at the upper end of this confidence interval, which represents an earnings gain of $43 per year, this impact on earnings would not be policy relevant.[58]

Exhibit 5-2.   Estimated Impacts of the Offset plus WIC on 2011-2015 Total Earnings and Total SSDI Benefits for Stage 1 Subjects

Outcome	T1 Mean	C1 Mean	Impact Estimate	Standard Error
Total earnings (January 2011 –December 2015)	$6,631	$6,622	$9[a]	$92
Total SSDI benefits (May 2011 –December 2015)	$54,155	$53,490	$665[###b]	$85

Source: SSA administrative records from the MEF and MBR.

Notes: Benefit outcomes are based on benefits due for the 2011-2015 period, corrected for retroactive adjustments made through May 2017 date. All earnings outcomes are based on a measure of earnings subject to Social Security taxes. See Appendix A for further details on outcomes. Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment. Means and impact estimates are regression-adjusted for baseline characteristics. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).

Unweighted sample sizes: T1 = 77,101; C1 = 891,429.

#/##/### Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a confirmatory standard of evidence (*p*-value adjusted by the multiple-comparisons procedure) and a two-tailed t-test with 9 degrees of freedom (resulting from a research design involving 10 study sites).

[a] The impact estimate for total earnings has a *p*-value after multiple-comparison adjustments of 0.927. Hence, the data do not provide confirmatory evidence of an impact.

[b] The impact estimate for total SSDI benefits has a *p*-value after multiple-comparison adjustment of 0.000. Hence, the data provide strong confirmatory evidence of an impact.

We do find strong confirmatory evidence of an impact of the BOND benefit offset on SSDI benefits. The estimated impact on benefits for the 56-month period is $665 per beneficiary ($143 per year or about $12 per month), which is slightly more than 1 percent of the control group average ($53,490). The 95 percent confidence interval is narrow and suggests that the effect lies between $101 and $184 per year (or between 0.9 and 1.6 percent of the control group average). We conclude that the offset policy did not decrease SSDI benefits.

---

[58]   The confidence interval may be interpreted as follows: if we replicated the demonstration many times under the same general conditions, and each time calculated an impact estimate and a 95 percent confidence interval with the same procedures, then the long-run average of the impact point estimates would be within the confidence interval in 95 percent of the replications.

5.2.2. Estimated Effects on Key Exploratory Outcomes

Although we do not find an effect on total earnings (the confirmatory outcome), the exploratory employment and earnings analysis suggests small behavioral work responses that are consistent with theory. Theory predicts increases in the percentage employed and the percentage with earnings above BYA and ambiguous effects on earnings above two or three times BYA.

A little more than a fifth (22.1 percent) of Stage 1 control group subjects worked at any point during the five-year follow-up period. Among the subjects who worked at any point, the average number of years in which some work occurred was three out of the five years. Consistent with the theory, there is evidence that the benefit offset increased the percentage employed and the percentage with earnings above BYA (Exhibit 5-3), but by relatively small amounts. The offset policy caused the percentage of beneficiaries with at least some employment during the five years to increase by 0.4 percentage points (a 2 percent increase relative to the control group employment rate of 22.1 percent). Likewise, the offset policy caused the share of beneficiaries whose earnings exceeded BYA in one or more years to increase by 0.4 percentage points from 5.1 to 5.5 percent, a 7 percent relative increase. There were also statistically significant increases in the number of years with employment and the number of years with earnings above BYA. All of these results point to small increases in earnings among those who would have earned less than BYA under current law.

In contrast, there is evidence of declines in earnings for those whose earnings would have been above BYA under current law, namely declines in the extent to which beneficiaries achieve earnings above two and three times BYA (Exhibit 5-3). The impact estimates for number of years with earnings in excess of two times BYA, number of years with earnings in excess of three times BYA, and the percentage with earnings above three times BYA in any year are each negative and statistically significant. However, the magnitudes of these effects are modest (6, 7, and 8 percent declines, respectively) relative to the control group averages. This evidence is consistent with the theoretical prediction that the offset will lead some subjects who would earn above BYA under current law to earn less under the offset.

These four behavioral effects—increases for any employment and for earnings above BYA, decreases for earnings above two and three times BYA—are each small in absolute and relative terms. Furthermore, they are counterbalancing, leading to a total impact on earnings that is essentially zero over five years.

Turning to benefit receipt, we estimate impacts consistent with the confirmatory evidence of a 1 percent increase in total SSDI benefits. The benefit offset increased the mean number of months with positive SSDI benefits by 0.8 months (a 2 percent increase relative to the control group average of 49.4 months). In addition, the percentage of Stage 1 treatment subjects with at least one month of SSDI benefits increased by 0.6 percentage points (a 1 percent increase relative to the control group percentage of 95.0).[59]

---

[59]   The reader may note that a small percentage of research subjects were not due any SSDI benefits during the analysis period. There are various reasons why some treatment and control group members never received SSDI benefits during the demonstration period. For example, SSA could have had suspended benefits because of SGA at the time the demonstration started and this information was updated retroactively after SSA drew the demonstration sample.

Above, we hypothesized that the SSDI benefit offset would decrease benefits under SSA's other disability program, SSI. However, we find no impact of the offset on SSI outcomes: total SSI payments due, the percentage of beneficiaries with at least one month with a SSI payment due, and the number of months with SSI payments due.

Exhibit 5-3.   Estimated Impacts of the Offset plus WIC on Other Employment Outcomes and Disability Benefits, 2011-2015, for Stage 1 Subjects

Outcome	T1 Mean	C1 Mean	Impact Estimate	Standard Error
Employment and Earnings Outcomes (January 2011 – December 2015)				
Employment during period (%)	22.47	22.11	0.36**	0.13
Number of years with employment	0.68	0.67	0.01*	<0.01
Earnings above BYA during at least one year (%)	5.51	5.13	0.38***	0.10
Number of years with earnings above BYA	0.13	0.13	0.01*	<0.01
Earnings above 2x BYA during at least one year (%)	2.15	2.23	-0.08	0.06
Number of years with earnings above 2x BYA	0.06	0.06	-0.00*	<0.01
Earnings above 3x BYA during at least one year (%)	1.14	1.22	-0.09*	0.04
Number of years with earnings above 3x BYA	0.03	0.03	-0.00*	<0.01
Disability Benefit Outcomes (May 2011 – December 2015)				
At least one month with SSDI benefits (%)	95.55	94.99	0.56***	0.09
Number of months with SSDI benefits	50.12	49.35	0.77***	0.07
Total SSI benefits	$1,945	$1,952	$-7	$34
At least one month with SSI benefits	19.49	19.59	-0.10	0.08
Number of months with SSI benefits	8.64	8.70	-0.06	0.05

Source: SSA administrative records from the MEF, MBR, and SSR.

Notes: SSDI and SSI benefit outcomes are based on benefits due for the 2011-2015 period, corrected for retroactive adjustments made through May 2017. All earnings outcomes are based on a measure of earnings subject to Social Security taxes. See Appendix A for further details on outcomes. Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment. Means and impact estimates are regression-adjusted for baseline characteristics. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).

Unweighted sample sizes: T1 = 77,101; C1 = 891,429.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test with 9 degrees of freedom (resulting from a research design involving 10 study sites) and with no multiple- comparisons adjustment.

5.2.3.   Stage 1 Effects By Year

Appendix F provides further evidence regarding how Stage 1 subjects responded to the benefit offset over time. Year-by-year results show two noteworthy trends. Importantly, several years of demonstration operations passed before small impacts on the share with earnings above BYA emerged. Impacts on the share with earnings above BYA were statistically significantly larger in both 2014 and 2015 than they were in 2012 (Exhibit F-1). In the final follow-up year of 2015, the estimate is about a quarter (0.27) of a percentage point (a 9 percent relative increase over the control group mean). Also, in 2015, a negative impact on the share with earnings above two times BYA emerged (with a magnitude double the size of the 2014 estimate, a statistically significant difference).

In addition, the SSDI benefits impacts grew larger over time. These impacts were statistically significantly larger in both 2014 and 2015 than they were in either 2012 or 2013 (Exhibit F-2).[60]

5.2.4. Stage 1 Subgroup Analysis

The Stage 1 subgroup analysis estimates impacts in seven pairs of subgroups.[61] For each pair, our primary interest is whether impacts on 13 outcomes differ between the two subgroups.[62] This analysis is considered exploratory. Given the large number of statistical tests involved (7 x 13 = 91), we expect to find, purely by chance, a least some differences in impact.

Taken together, the following four points suggest that the subgroup analysis of Stage 1 adds little of substantive interest to the analysis of the full sample.

First, none of the seven subgroup pairs show evidence of a discernable difference in the effect of the benefit offset on total earnings, one of the confirmatory outcomes.[63] Furthermore, as is found for the whole sample, none of the individual subgroups have an effect on total earnings.

Second, in the one subgroup pair that shows a pattern of differences in impacts—defined by employment in 2010—the magnitudes are consistent with approximately equal proportional effects (with employment in 2010 leading to larger absolute impacts). Those subjects who had employment in 2010 have larger behavioral effects in absolute terms than those who did not work in 2010 (Exhibit F-20). Impacts for those employed in 2010 are significantly more positive for employment, any year with earnings above BYA, and number of years with earnings above BYA. However, although the impacts for those employed in 2010 are larger in absolute terms, their magnitudes in relation to the control group averages are similar

---

[60]   Corresponding to the increasing impact on yearly SSDI benefits, the impact on the number of months with SSDI receipt statistically significantly increased each year from 2012 to 2015. The impact on any SSDI receipt also increased over time with the 2013 and 2014 impacts larger than the 2012 impact and the 2015 impact larger than the 2012, 2013, and 2014 impacts.

[61]   The seven subgroup pairs are: short-duration SSDI receipt compared to long-duration SSDI receipt; SSDI-only beneficiaries compared to concurrent beneficiaries; employed in 2010 compared to not employed in 2010; access to Medicaid buy-in programs compared to no access to Medicaid buy-in programs; age 49 or less at baseline compared to age 50 or more at baseline; primary impairment of major affective disorder compared to all other primary impairments; and primary impairment of back disorder compared to all other primary impairments.

[62]   These 13 outcomes include all the confirmatory and key exploratory outcomes in Exhibits 5-2 and 5-3, with the exceptions of any SSDI benefits due and any SSI benefits due.

[63]   Differences found in two subgroup pairs for the effect of the benefit offset on the other confirmatory outcome, SSDI benefits, appear to be driven by the windfall (non-behavioral) effect of the offset. Thus, when one subgroup has a higher level of employment and earnings above BYA than its complementary subgroup, we expect the mechanical windfall effect of the offset to be greater in this subgroup. This results in a larger positive effect on SSDI benefits, in the absence of a difference in behavioral response. The analysis considers differences in behavioral effects as more noteworthy than differences that stem from the mechanical windfall effect of the offset.

to those found among subjects not employed in 2010.[64] Those with employment in 2010 also have larger impacts on SSDI benefits, and months with SSDI benefit receipt. These larger impacts appear to be the result of differences between the subgroups in average number of years earning above BYA, rather than differences in behavioral effects.[65]

Third, even where theory suggested differences might emerge, differences were not detected. We had expected behavioral responses to the benefit offset policy to be larger for: those beneficiaries who had received SSDI benefits for 36 or fewer months at random assignment relative to those who received them for at least 37 months (Exhibit F-18)[66]; SSDI-only beneficiaries relative to those also concurrently receiving SSI (Exhibit F-19); those residing in states with Medicaid Buy-In (MBI) programs relative to those residing in other states (Exhibit F-21)[67]; and those under age 50 at baseline relative to those ages 50 to 59 at baseline (Exhibit F-22).[68]

And fourth, no differences in impacts were found for those with major affective disorder compared to all other impairments (Exhibit F-23), or for those with back disorders compared to all other impairments (Exhibit F-24).

5.2.5.   Stage 1 Effects on Non-Countable Earnings

We find evidence that the offset policy decreases "non-countable earnings"—dollar amounts recorded in the SSDI benefit payment system that reduce the earnings amount use for benefit calculation. Non-countable earnings are principally comprised of impairment-related work expenses, subsidies, and special condition amounts (Exhibit 5-4).[69]

---

[64]   For example, the impact for those employed in 2010 on the share with earnings above BYA is 6 percent of the subgroup's control group share. The impact on this outcome for those without employment in 2010 is smaller in absolute terms but still 9 percent of the share among the control subjects in the subgroup.

[65]   Because the employed in 2010 subgroup has more years earning above BYA during the follow-up period than no employment in 2010 subgroup, the windfall (non-behavioral) increase in SSDI benefits is greater for the employed in 2010 subgroup.

[66]   Although impacts were significantly more negative for the two times BYA outcomes in the short-duration subgroup no other evidence was found for differences in behavioral effects.

[67]   The impact on employment was larger for those residing in states with MBI programs. Although that single result may be due to the availability of an MBI program, it might also be due to other characteristics that distinguish these areas. If the availability of the MBI interacted with the size of impacts for the offset, we would expect the evidence to emerge in the form of differences in impacts on the share with earnings in the range above BYA, the earnings range over which the offset is relevant. No statistically significant differences emerged for these outcomes.

[68]   The analysis found larger positive impacts on the SSDI benefits due outcomes for those under age 50 at baseline relative to those ages 50 to 59. The apparent reason for the larger benefit impacts is the greater average number of years with earnings above BYA in the younger subgroup, creating a larger windfall effect for those under age 50 at baseline.

[69]   "Subsidies" and "special conditions" refer to support and on the job assistance provided by an employer or outside agency that may result in a beneficiary receiving more pay than the actual value of services performed (SSA 2017d).

Non-countable earnings decrease the earnings amount SSA uses when determining SSDI benefit amounts under both the BOND offset and current law. For beneficiaries subject to current law, non-countable earnings can reduce earnings counted from above the SGA threshold to below the threshold, and thus make the difference between receiving and forgoing a monthly benefit check.

Relative to current law, the benefit offset reduces the incentive to report non-countable earnings. This is because, for BOND treatment group beneficiaries with earnings in excess of BYA, every $2 in non-countable earnings reported will result in $1 more in benefits. Therefore, under the offset rules, non-countable earnings only make a difference between larger and smaller partial benefits (rather than the difference between full and zero benefits). Consistent with these contrasting incentives, we find that the benefit offset decreases the proportion of T1 subjects with non-countable earnings by 0.29 percentage points, a decrease of about a third relative to the C1 average of 0.81 percent. We also find a negative impact on the amount of non-countable earnings, driven largely by the decrease in the percentage of subjects who report any non-countable earnings.[70]

Exhibit 5-4.   Impacts on Non-Countable Earnings: T1 Versus C1

Outcome	T1 Mean	C1 Mean	Impact Estimate
Any non-countable earnings (May 2011-December 2015) (%)	0.53	0.81	-0.29*** (0.04)
Amount of non-countable earnings (May 2011-December 2015)	$33	$56	-$23*** ($4)

Source: SSA administrative records from the DAF (2015) and MBR.

Notes: Non-countable earnings include impairment-related work expenses, subsidies, and special conditions. Non-countable earnings also include sick and vacation pay, but these are not included in our estimates because they are not available in the data we used for this analysis (the DAF). Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).

Unweighted sample sizes: T1 = 77,101; C1 = 891,429.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test with 9 degrees of freedom (resulting from a research design involving 10 study sites) and with no multiple- comparisons adjustment.

5.3.   Stage 2

5.3.1.   Estimated Effects on Confirmatory Outcomes

The Stage 2 analysis finds no confirmatory evidence of an impact on total earnings during the 2012 to 2015 period for either Stage 2 treatment group relative to current law (Exhibit 5-5). For both T21 and T22 groups, the point estimates for impacts on earnings are positive and moderately-sized relative to the control group average (9 and 10 percent, respectively). The impact estimate magnitudes are relatively close to statistical significance at the study's threshold of 0.10, but are not statistically significant. The 95 percent confidence interval for T21 versus C2 ranges from $2 to $748 per year, or from 0 to 18 percent of

---

[70]   The negative impact on the amount of non-countable earnings across all Stage 1 subjects is equal to 41 percent of the control group average. This negative impact is produced by a 35 percent relative drop in the share who report non-countable earnings and a 10 percent relative drop in reported amount conditional on reporting.

the control group average.[71] For T22 versus C2, the confidence interval ranges from -$214 to +$1,016 per year, or from -5 to +25 percent of the control group average.

Exhibit 5-5.   Estimated Impacts of the Offset plus WIC or EWIC on 2012-2015 Earnings and SSDI Benefits for Stage 2 Volunteers: All Policy Comparisons

Outcome	Average Outcome with Offset and WIC (T21)	Average Outcome with Offset and EWIC (T22)	Average Outcome under Current Law (C2)	Estimated Impact of Offset + WIC vs Current Law (T21 vs. C2)	Estimated Impact of Offset + EWIC vs Current Law (T22 vs. C2)	Estimated Impact of EWIC instead of WIC Given Offset (T22 vs. T21)
Total earnings (January 2012–December 2015)	$18,009	$18,115	$16,510	$1,499[a] ($660)	$1,605[a] ($1,088)	$106 ($726)
Total SSDI benefits (January 2012–December 2015)	$51,423	$51,630	$49,633	$1,791[##b] ($463)	$1,997[##b] ($529)	$207 ($614)

Source: SSA administrative records from the MEF and MBR and the Stage 2 Baseline Survey.

Notes: SSDI and SSI benefit outcomes are based on benefits due during the 2011-2015 period, corrected for retroactive adjustments made through May 2017. All earnings outcomes are based on a measure of earnings subject to Social Security taxes. See Appendix A for further details on outcomes. Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national population of SSDI-only beneficiaries who would volunteer for study enrollment. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).

Unweighted sample sizes: T21 = 4,854, T22 = 3,041, C2 = 4,849.

#/##/### Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a confirmatory standard of evidence (*p*-value adjusted by the multiple-comparisons procedure) and a two-tailed t-test with 9 degrees of freedom (resulting from a research design involving 10 study sites).

[a] The impact estimates for total earnings for T21 vs. C2 and for T22 vs. C2 had *p*-values after multiple-comparison adjustments of 0.108 and 0.194, respectively. Hence, the data do not provide confirmatory evidence of an impact in either case. Readers may note that the t-statistics derived by dividing the T21 vs. C2 and T22 vs. C2 estimates by their respective standard errors are 2.3 and 1.5, respectively. In the absence of adjustments for multiple comparisons and degrees of freedom, the respective p-values would be 0.049 and 0.174, making the T21 vs. C2 estimate significant at the 0.05 level. Both adjustments have substantial effects on p-values, however. The degrees of freedom adjustment is substantial because it is based on the number of sites, which is small (10).

[b] The impact estimates for total SSDI benefits for T21 vs. C2 and for T22 vs. C2 both had *p*-values after multiple-comparison adjustments of 0.020. Hence, the data provide confirmatory evidence of an impact.

We find confirmatory evidence in Stage 2 that relative to current law, the offset policy increases SSDI benefits (Exhibit 5-5). For the four-year period, the impact estimates for T21 and T22 relative to the control group are $1,791 and $1,997, respectively (both are about 4 percent of the control group mean). These impacts represent increases of $37 and $42 per month and about $450 to $500 per year relative to the SSDI benefits of the C2 group. The 95 percent confidence intervals for these impacts are from $186 to $709 per year (T21 versus C2) and from $200 to $799 per year (T22 versus C2).

We find no evidence that the more intensive EWIC services delivered to T22 subjects, relative to the WIC services offered to T21 subjects, affected earnings or SSDI benefits (Exhibit 5-5).

---

[71]   The reader may note that the confidence interval does not include $0. Even though this is the case, the impact estimate fails to achieve statistical significance because of the multiple comparisons adjustment performed for this confirmatory outcome.

5.3.2. Estimated Effects on Key Exploratory Outcomes

As with Stage 1, the Stage 2 exploratory impact estimates provide evidence of some behavioral responses to the benefit offset policy. As predicted by theory, we find the benefit offset increases both employment and the percentage of subjects with earnings above BYA (Exhibit 5-6).

Exhibit 5-6. Estimated Impacts of the Offset plus WIC or EWIC on Other Employment Outcomes and Disability Benefits, 2012-2015, for Stage 2 Volunteers: All Policy Comparisons

Outcome	Average Outcome with Offset and WIC (T21) (1)	Average Outcome with Offset and EWIC (T22) (2)	Average Outcome under Current Law (C2) (3)	Estimated Impact of Offset + WIC vs Current Law (T21 vs. C2) (4)	Estimated Impact of Offset + EWIC vs Current Law (T22 vs. C2) (5)	Estimated Impact of EWIC instead of WIC Given Offset (T22 vs. T21) (6)
Employment and Earnings Outcomes (January 2012 –December 2015)						
Employment during period (%)	54.53	55.48	52.49	2.04* (0.95)	2.99* (1.38)	0.95 (1.10)
Number of years with employment	1.56	1.56	1.48	0.08** (0.03)	0.07* (0.03)	-0.01 (0.03)
Earnings above BYA during at least one year (%)	19.97	19.31	15.99	3.98*** (0.93)	3.32*** (0.89)	-0.67 (0.91)
Number of years with earnings above BYA	0.46	0.46	0.37	0.09*** (0.02)	0.09*** (0.02)	0.00 (0.02)
Earnings above 2x BYA during at least one year (%)	7.64	6.88	6.71	0.93 (0.57)	0.17 (0.59)	-0.76 (0.61)
Number of years with earnings above 2x BYA	0.17	0.15	0.15	0.02 (0.01)	0.01 (0.01)	-0.02 (0.02)
Earnings above 3x BYA during at least one year (%)	3.22	3.17	3.06	0.16 (0.41)	0.11 (0.49)	-0.05 (0.41)
Number of years with earnings above 3x BYA	0.07	0.07	0.07	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
Disability Benefit Outcomes (January 2012 –December 2015)						
At least one month with SSDI benefits (%)	97.64	96.97	96.53	1.11** (0.39)	0.45 (0.46)	-0.66 (0.44)
Number of months with SSDI benefits	44.53	44.48	42.40	2.13*** (0.25)	2.08*** (0.28)	-0.05 (0.28)
Total SSI benefits	$143	$139	$129	$13 ($33)	$10 ($36)	$-4 ($43)
At least one month with SSI benefits (%)	2.98	2.96	2.92	0.07 (0.39)	0.04 (0.47)	-0.03 (0.52)
Number of months with SSI benefits	0.75	0.81	0.71	0.05 (0.11)	0.10 (0.15)	0.06 (0.18)

Source: SSA administrative records from the MEF, MBR, and SSR and the Stage 2 Baseline Survey.

Notes: Benefit outcomes are based on benefits due during the 2012-2015 period, corrected for retroactive adjustments made through May 2017. All earnings outcomes are based on a measure of earnings subject to Social Security taxes. See Appendix A for further details on outcomes. Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national population of SSDI-only beneficiaries who would volunteer for study enrollment. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).

Unweighted sample sizes: T21 = 4,854, T22 = 3,041, C2 = 4,849.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test with 9 degrees of freedom (resulting from a research design involving 10 study sites) and with no multiple- comparisons adjustment.

Compared to Stage 1 subjects, Stage 2 volunteers in the treatment and control groups were much more likely to have worked during the follow-up period. A little more than half of Stage 2 current-law subjects (52 percent) were employed at some point during the four-year follow-up, compared to 22 percent of Stage 1 current-law subjects within five years. Exhibit 5-5 shows positive impacts of the offset on the proportions of those with any employment during the follow-up period and the number of years employed during that period. The shares of beneficiaries with employment in at least one year are 54.5 percent for the T21 group and 55.5 percent for the T22 group compared to 52.5 percent for the C2 group (representing increases of 4 and 6 percent over the current law proportion).

We also find substantive impacts in Stage 2 of the benefit offset policy on the proportion earning more than the BYA amount for at least one year. The impact of the benefit offset on the share with earnings above BYA for one or more years is 4.0 percentage points for T21 and 3.3 percentage points for T22 (representing 25 and 21 percent, respectively, of the control group proportion of 16.0 percent).

For Stage 2, we find no evidence of impacts of the offset interventions versus current law on proportions of beneficiaries with earnings above two times or three times BYA.

The Stage 2 benefit-related results show positive impacts of the offset policy on SSDI benefit-related outcomes. Complementing the confirmatory outcome result of a 4 percent increase in SSDI benefits, the offset increases the number of months with SSDI benefit payments due by about 2 months for both the T21 and T22 groups relative to the control group (both are 5 percent of the control group mean). For T21, we also found a positive impact on the percentage having any SSDI benefits (a one percent increase relative to the C2 mean).

Similar to Stage 1, all Stage 2 SSI related outcomes have no detectible behavioral response to the benefit offset.

Like the Stage 2 confirmatory outcomes, for all the Stage 2 exploratory outcomes presented in this chapter, there are no statistically significant differences in outcomes between T21 and T22 subjects. Thus, for a robust set of earnings and benefit-related outcomes, there are no differences in the behavioral response of volunteers to the demonstration based on whether they were offered WIC services or the more intensive EWIC services.[72]

5.3.3. Stage 2 Effects by Year

In Appendix F we present estimates for annual impacts in Stage 2 on the earnings and benefit variables (Exhibits F-7 through F-9). There were no notable trends in impacts on earnings-related outcomes.[73] In

---

[72]   We performed an additional exploratory analysis that pools the two treatment groups in order to estimate an average effect of the benefit offset versus current law in Stage 2. The results from this non-pre-specified analysis are presented and discussed in Appendix F (Exhibit F-10).

[73]   We tested whether impact estimates were equal within three pairs of years: 2013 and 2014; 2013 and 2015; and 2014 and 2015. (The years of 2013, 2014, and 2015 are the calendar years that occur entirely after the end of Stage 2 random assignment.) Out of 45 tests (three policy comparisons, five outcomes, three pairs of years), only one pair of impact estimates on an earnings-related outcome was statistically significant: for the T22 versus T21 policy contrast the impact on employment increased from 2013 to 2015. However, this result may

contrast, the impacts on SSDI benefits, number of months of SSDI benefits, and percentage with any SSDI benefit receipt all increase over time for the T21 versus C2 comparison. Impacts on the latter two outcomes also increase over time for the T22 versus C2 comparison.[74] There were no statistically significant trends for any benefit-related outcomes in the T22 versus T21 comparison.

5.3.4. Stage 2 Subgroup Analysis

We present impact estimates for seven pairs of Stage 2 subgroups in Appendix F (Exhibit F-25 through F-52). For each pair, three comparisons provide evidence about the effects of the benefit offset versus current law: T21 versus C2, T22 versus C2, and the average of T21 plus T22 versus C2 (with the pooling of T21 and T22 intended to increase statistical power). As in the full sample analysis, the T22 versus T21 comparison provides evidence about the effects of EWIC relative to WIC for subjects offered the benefit offset. A large number of statistical tests are performed in the Stage 2 subgroup analysis (7 subgroup pairs x 4 comparisons x 13 outcomes = 364 tests of differences in impacts). Hence, we expect a substantial number of findings of statistical significance purely by chance. We make no adjustment to statistical significance based on the large number of tests. Therefore, even statistically significant results should be interpreted with caution.

We find no clear pattern of evidence that any subgroup has stronger behavioral effects in either the offset-versus-current law comparisons or in the EWIC-versus-WIC comparison.

First, for the offset versus current law, we find inconsistent evidence of subgroup differences in behavioral effects in the employed at baseline and not employed at baseline subgroups.[75] The not employed at baseline subgroup has a larger positive effect on employment, in both absolute and relative terms, than the employed at baseline subgroup. But those employed at baseline have larger absolute impacts on percentage with earnings above BYA in one or more years and number of years with earnings above BYA. We interpret this evidence as showing similar behavioral effects in the two subgroups[76], with the statistically significant differences in impacts resulting from the very different starting places for the two subgroups: being non-employed at baseline left great scope for an impact on the employment rate while being employed at baseline positioned more beneficiaries to increase their earnings above BYA. The greater positive impact on SSDI benefits for those employed at baseline is also likely an artifact of different levels of earnings between the subgroups (leading to a larger windfall effect for those employed at baseline), rather than a difference in effects on earnings behavior.

---

be spurious, as there are no statistically significant overall impacts of T22 versus T21 for employment in any year.

[74] For T21 versus C2, the impacts on SSDI benefits and months with SSDI benefits are higher in 2015 than in the two prior years. For T22 versus C2, the impact on months with SSDI benefits is higher in 2015 and 2014 than in 2013.For both T21 versus C2 and T22 versus C2, from 2013 to 2015, the impact on the percentage of beneficiaries with any SSDI benefits grows statistically significantly larger each year.

[75] See Exhibits F-29, F-30, and F-32.

[76] The similar effects are seen in the relative-to-control-mean sizes of the impacts on the BYA outcomes. The impacts on any year earning above BYA for the two subgroups are 20 percent (for employed at baseline) and 31 percent (for not employed at baseline) increases relative to their respective control means. The impacts on number of years earning above BYA are 22 percent and 29 percent relative to their respective control means.

Second, three other subgroup pairs offer only sparse evidence of differences in impacts between subgroups. One subgroup pair offers plausible, though weak, evidence of larger behavioral effects of the benefit offset. Effects on employment and earnings above BYA are larger for those with educational attainment of less than an associate's degree than for those with a postsecondary degree in 2 of the 12 tests related to these outcomes (Exhibits F-49, F-50, and F-52). Because employment is lower in the subgroup with lower educational attainment, perhaps the potential for the benefit offset to affect behavior was greater in this subgroup.

Another subgroup pair offers weak evidence of larger behavioral effects of EWIC relative to WIC. Those beneficiaries with the primary impairment of major affective disorder exhibit larger impacts of EWIC on earning over BYA than beneficiaries with all other impairments on both of the tests related to this outcome (Exhibit F-43). It is plausible that the more extensive counseling contacts offered by EWIC helped beneficiaries with major affective disorders more than other beneficiaries.

Finally, Exhibits F-45 to F-48 show some unexpected impact differences between the subgroup of beneficiaries with the primary impairment of back disorder relative to beneficiaries with all other impairments. Within the back disorder subgroup, there are more negative effects on earnings measures for the offset-plus-EWIC group compared to current law and compared to the offset-plus-WIC group. These results appear to be driven by the low average earnings and employment of the T22 subjects with the primary impairment of back disorder. This subgroup of T22 subjects is the smallest of the 14 subgroups examined (sample size = 424 subjects). It is not clear why the offset-plus-EWIC combination would depress earnings compared to current law and the offset plus WIC for those with back disorders.

Other scattered findings of statistically significant differences in impacts appear to be due to windfall (i.e., non-behavioral) effects or seem implausible. The subgroup pairs with such results include: beneficiaries under age 50 at baseline relative to beneficiaries ages 50 and above[77] and beneficiaries residing in states with MBI programs relative to beneficiaries residing in other states. Given the concerns about multiple statistical tests, these results do not seem to warrant extended discussion. No statistically significant differences in impacts are found for subgroups defined by duration of SSDI receipt.

5.3.5. Stage 2 Effects on Non-Countable Earnings

Finally, as in Stage 1, we find evidence that the offset policy decreases the non-countable earnings that beneficiaries report to SSA (Exhibit 5-7). We find that the share reporting non-countable earnings in the T21 group is about one-third less than the C2 group rate of 3.9 percent. The average amount of non-countable earnings across all subjects (including those not reporting any non-countable earnings) is lower in the T21 and T22 groups than it is in the C2 group. This is consistent with an explanation that possible suspension of SSDI benefits creates a more compelling reason to report non-countable earnings than the $1 for $2 offset. We do not detect an effect of EWIC versus WIC on either the percentage reporting non-countable earnings or the average amount of non-countable earnings.

---

[77]   The larger positive impact on number of months with SSDI payments due for those under age 50 (on 2 of the 3 tests related to this outcome) appears to arise because of the higher number of years with earnings above BYA during follow-up in the younger subgroup (Exhibits F-37, F-38, and F-40). The higher number of years with earnings above BYA leads to a larger windfall effect for those under age 50.

Exhibit 5-7.   Impacts on Non-Countable Earnings for Stage 2 Volunteers: All Policy Comparisons

Outcome	Average Outcome with Offset and WIC (T21) (1)	Average Outcome with Offset and EWIC (T22) (2)	Average Outcome under Current Law (C2) (3)	Estimated Impact of Offset + WIC vs Current Law (T21 vs. C2) (4)	Estimated Impact of Offset + EWIC vs Current Law (T22 vs. C2) (5)	Estimated Impact of EWIC instead of WIC Given Offset (T22 vs. T21) (6)
Any non-countable earnings (January 2012 - December 2015)	2.65	2.94	3.90	-1.18** (0.42)	-0.87 (0.48)	0.31 (0.44)
Amount of non-countable earnings (January 2012 - December 2015)	$131	$73	$230	$-99* ($54)	$-157*** ($46)	$-58 ($35)

Source: SSA administrative records from the DAF (2015) and MBR and the Stage 2 Baseline Survey.

Notes: Non-countable earnings include impairment-related work expenses, subsidies, and special conditions. Non-countable earnings also include sick and vacation pay, but these are not included in our estimates because they are not available in the data we used for this analysis (the DAF). Weights are used to ensure that the BOND subjects who met analysis criteria are representative of the national population of SSDI-only beneficiaries who would volunteer for study enrollment. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).

Unweighted sample sizes: T21 = 4,854 ,T22 = 3,041 , C2 = 4,849

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test with 9 degrees of freedom (resulting from a research design involving 10 study sites) and with no multiple-comparisons adjustment.

5.4.   Impacts on Survey-Derived Outcomes in Stages 1 and 2

The impact analysis also examined a large number of outcomes constructed from Stage 1 and Stage 2 survey data. These impact results are presented in Hoffman et al. (2017), Gubits et al. (2017), and Geyer et al. (2018b).

No meaningful effects on survey-measured outcomes were found for Stage 1 beneficiaries. Neither did the analysis find meaningful effects in the comparison of EWIC to WIC in Stage 2.

There were some meaningful impacts for Stage 2 treatment subjects (pooled T21 and T22) relative to control group subjects. Survey analysis showed that the offset plus WIC, compared to current law, led to increased work hours and job search (when not working), more jobs with paid vacation, fewer households living below the federal poverty line, and smaller amounts of income support received from irregular sources outside the household. There was also evidence that the offset combined with EWIC had additional effects on beneficiaries compared to current law. These effects included increased work hours and job search (when not working), increased personal goal-setting among beneficiaries to stop receiving SSDI, and reports that fewer households were living below the federal poverty line.

5.5.   Discussion

For both Stage 1 and Stage 2 of BOND, we find confirmatory evidence that the benefit offset led to higher SSDI benefits relative to current law. In contrast, for neither Stage 1 nor Stage 2 do we find confirmatory evidence of an effect of the benefit offset on total earnings.

For Stage 1, we find an essentially zero net effect on total earnings over the five-year follow-up period (2011 to 2015). This effect is precisely estimated and the 95 percent confidence interval does not include any policy relevant impact. Underlying this essentially zero net effect, the analysis detects small,

theoretically-predicted behavioral responses to the benefit offset policy in opposite directions. Some beneficiaries who would have earned less than BYA under current law increased their earnings because of the benefit offset (seen in the positive effect on the proportion with at least one year of BYA-level earnings), while other beneficiaries who would have earned above BYA under current law decreased their earnings (seen in the negative effect on the proportion with at least one year of 3x BYA-level earnings). The opposing effects nearly cancel each other out resulting in the measured impact on average total earnings that is essentially zero (less than 0.2 percent).[78]

For Stage 1, the offset increased SSDI benefits by an average of $665 over 56 months, or $12 per month (1 percent of the control group mean). The positive sign of this impact shows that the expansion of the number of treatment group subjects with at least one year of earnings above BYA—beneficiaries who should have received partial benefits under the offset rather than full benefits—did not produce sufficient savings to reduce total benefits. In other words, it did not outweigh the increase in benefits for subjects earning more than the BYA amount who became eligible for a partial benefit under the offset rules. The increase (i.e., the impact estimate) in the percentage with at least one year of above-BYA earnings is only about four-tenths of a percent of the sample compared to the 5.1 percent of the control group with at least one year of above-BYA earnings. The apparent reduction in earnings of the above BYA group in response to the offset served to increase their SSDI benefits beyond the mechanical windfall effect.

The Stage 2 results show that beneficiaries who volunteered for the demonstration experienced impacts on employment and earnings above BYA larger than the typical Stage 1 subject.[79] We find an effect of the benefit offset on total earnings has a 95 percent confidence interval that ranges from -2 percent to +20 percent of the control group average. Because the level of earnings in the control group is low, the upper bound of the 95 percent confidence interval implies an earnings increase of $127 per month for Stage 2 treatment subjects with any employment during the follow-up period.[80]

As in Stage 1, the statistically insignificant effect of the offset on earnings in Stage 2 was not enough to produce a reduction in total SSDI benefits. Rather, the average amount of SSDI benefits due to the Stage 2 treatment subjects was about 4 percent larger than the amount due to Stage 2 control subjects.

We find no evidence that enhancements to counseling services under EWIC had marginal effects on the offset's impact on earnings or benefit outcomes. Thus, although enhanced counseling had small, positive impacts on understanding of the offset (Chapter 3) and T22 subjects who used the offset experienced shorter adjustment times and fewer overpayments compared to T21 subjects (Chapter 4), these improved process outcomes do not translate into differences in earnings or benefits. The results in this section are consistent with the finding in Chapter 4 that EWIC services did not increase offset use compared to WIC services.

---

[78] The Stage 1 estimated impact on total earnings (2011 to 2015) of $9 is less than 0.2 percent of the C1 group mean of $6,622.

[79] A test for difference across the two stages in size of effect on employment is statistically significant at the 0.10 level. The difference in effect on the percentage earning above BYA is statistically significant at the 0.01 level.

[80] The upper bound of the confidence interval is $3,335 over four years for the whole sample, or $69 per month. The $127 amount is produced by dividing $69 per month by the 55 percent of Stage 2 treatment subjects with any employment during the follow-up period.

In a few instances, this chapter has made comparisons between Stage 1 and Stage 2 impact estimates. Chapter 7 goes further, using findings from the two stages to make inferences about the impacts for Stage 1 SSDI-only treatment subjects who, if solicited, would not have volunteered for Stage 2. The findings shed additional light on the interpretation of the estimates presented here.

6.   The Benefit-Cost Analysis of BOND

The previous chapter examined the impacts of the BOND benefit offset, including the effects of the benefit offset on earnings and SSDI payments. This chapter considers whether the various impacts of the BOND interventions—when viewed in aggregate—resulted in overall gains or losses to beneficiaries, to SSA's Disability Insurance Trust Fund, and to society as a whole.

In brief, the per-beneficiary findings of this benefit-cost analysis are as follows. Applying the benefit offset intervention to the full range of SSDI beneficiaries creates small gains for beneficiaries and small costs for the Trust Fund—both effects primarily resulting from increased SSDI benefit payments. These effects create a slight loss for society as a whole. Within the small subset of beneficiaries who volunteered for the offset in Stage 2 of the demonstration, the benefit offset plus WIC produced moderate gains per beneficiary and an overall societal gain. Although the offset plus EWIC helped volunteers slightly more than the offset plus WIC, it imposed much greater costs on the Trust Fund and produced an overall societal loss, mainly due to higher counseling costs.

The chapter begins with an explanation of how the benefit-cost measures are derived from BOND evaluation data and from external sources. Findings for Stage 1 follow, first the "base case" findings produced by preferred analytic assumptions and then alternative results showing the main findings to be robust to most alternative scenarios. Similar sections for the Stage 2 benefit-cost analysis come next. The final section summarizes the findings.

6.1.   Methods Used in the Benefit-Cost Analysis

The benefit-cost analysis considers a wide variety of benefits and costs. The benefits and costs are summed up from four different perspectives, those of:

- SSDI beneficiaries,

- The government's Disability Insurance Trust Fund,

- The rest of the government, and

- Society as a whole,

where society as a whole combines the other three perspectives plus non-beneficiaries who may have been affected. Unless specifically noted, all the findings presented in this chapter, like those appearing elsewhere in this report, are reported on a per beneficiary basis. Because a large percentage of the treatment groups did not use the benefit offset, per beneficiary benefits and costs reflect averages for a population that includes many individuals whose benefits and costs were unaffected by the offset policy. For example, the cost of increases in SSDI benefits to the DI Trust Fund is simply an average of the change in SSDI benefits among those whose benefits changed and zero for the much larger set of beneficiaries for whom there was no change.

The impact analysis described in the previous chapter provides estimates of some of the needed benefit and cost measures. Appendix G collects the relevant estimates in one place. It also provides details of the benefit-cost analysis methodology not covered in the text. The estimates of impact that appear in Exhibits G-2 and G-5 provide, for Stage 1 and Stage 2 respectively, the best quantitative information available

about the true effects of the benefit offset and enhanced work incentives counseling. For example, the positive estimates of impact for Stage 1 earnings in 2013 through 2015 imply that true impacts were more likely to be positive than negative during those years.[81] Conversely, the negative estimates for 2011 and 2012 imply the opposite.

Impact analysis places greatest emphasis on obtaining highly confident (i.e., statistically significant) evidence that effects differ from zero. Benefit-cost analysis instead uses "best guess" indications of effect magnitudes regardless of the degree of confidence (although statements of confidence, in Bayesian statistical terms, are made). Thus, even when point estimates are not statistically significant, we use them in the benefit cost analysis. The lack of statistical significance for these findings does not indicate that each of the true impacts is zero, although a value of zero is a possibility. Instead, the lack of statistical significance for these findings means that the evidence is inconclusive regarding whether a non-zero effect has occurred. In this situation the measure of impacts in Exhibits G-2 and G-5, almost all of them not zero, provide the best available information on true impact magnitude. These figures are used, regardless of statistical significance.

6.1.1. Indirectly Determined Benefits and Costs

Several key impact estimates listed in Appendix G—the effects of the demonstration's intervention on earnings, SSDI benefits, and SSI benefits—feed into measurement of other benefits and costs used in the analysis. For example, an increase in earnings engenders changes in fringe benefits, work-related expenditures, and time available outside of work (time that is of value to most persons). Changes to these outcomes were not measured directly by the BOND evaluation. Instead, as is standard in benefit-cost analyses, we use external (i.e., non-BOND evaluation) estimates from various sources to determine how these benefits and costs change as earnings change. For example, the increase in the value of fringe benefits attributable to a one dollar increase in earnings is derived from an externally-obtained multiplier. Appendix G describes the sources used to obtain each of the multipliers used to translate measured earnings effects into benefit and cost impacts for other outcomes.

Increases in both earnings and SSDI payments can cause increases in income taxes. But the ramifications here are not traced by a single multiplier applied to every dollar of income. For several reasons, the relationships are more complex:

- Taxes are a nonlinear function of income (so no single multiplier is appropriate);

- Income taxes are dependent on the type of income received (for example, federal income taxes treat SSDI benefits and earnings differently); and

---

[81]   The statements of likelihood in this paragraph arise from Bayesian statistical theory which, unlike frequentist statistical theory, combines initially agnostic "prior" beliefs about impact magnitudes with the evidence of the data to form "posterior" beliefs about the chances that true impacts of various magnitudes exist. From a posterior distribution one can state the likelihood that a given true impact is positive and the likelihood that it is negative. Making the usual assumption that the data follow a normal distribution together with an agnostic prior, the posterior distribution of any cost or benefit item is symmetric around the measured level of that factor in the data, placing equal shares of the likelihood on either side. See Winkler (2003).

- Changes in income are heterogeneous (for example, in Stage 1, there are positive impacts in 2015 on the prevalence of earnings above BYA but negative impacts on the prevalence of earnings above twice BYA).

For these reasons, we estimate increases in payroll and federal and state income taxes through *IncTaxCalc*, a tax calculator program (Bakija 2016) that translates income amounts into tax liabilities (see Appendix G). Stage 1 income used for this purpose includes only earnings and SSDI and SSI benefits—the only sources available for the full sample given the limited sample coverage of the Stage 1 36-month follow-up survey. For comparability of findings across stages, the same inputs are used in Stage 2.[82]

As detailed in Appendix G, after assuming that all increases in net income were spent—and likewise that all decreases in income result in decreased spending—the calculation of sales taxes paid uses the average state and local sales tax percentage for the U.S. derived from data available from the Tax Foundation.

Using federal government administrative data, the benefit-cost study also includes estimates of the offset's impacts on costs incurred under the Ticket-to-Work (TTW) Program and in providing state Vocational Rehabilitation (VR) services. Costs of administering the offset policy relative to costs of administering current law are based upon federal administrative data and data provided by the BOND implementation team. Costs of work incentives counseling are based upon data provided by the implementation team.

6.1.2. Arriving at a "Bottom Line" of Net Benefits from a Variety of Perspectives

For benefits and costs measured in dollars, summation across all dollar-denominated benefits and costs tells us whether gains due to the BOND interventions exceeded losses. However, sufficient information does not exist to value every potential impact of the offset and work incentives counseling in dollars. For example, studied interventions may have affected the health and life-satisfaction of beneficiaries. Moreover, the well-being of non-beneficiaries may have been affected because the interventions induced more beneficiaries to work. On the one hand, this may have conveyed psychic benefits to non-beneficiaries who want more income-transfer recipients to work; on the other hand, it may have displaced non-beneficiaries from jobs. To the extent possible, the analysis judges whether, if they could actually be measured in dollars, these potential effects would be likely to overturn the conclusions based solely on the monetized benefits and costs.

A further potential effect on non-beneficiaries—one estimated in dollars in the benefit-cost analyses—is deadweight loss. Deadweight loss results if the tested interventions affect the government's fiscal position—for example, by causing government to pay a larger amount of SSDI benefits. The assumption is that when government outlays change, some government-imposed taxes at some point in time must commensurately grow larger or smaller than they otherwise would be.[83] Taxes create economic

---

[82]   The Stage 2 36-month follow-up survey encompasses most of the sample for that stage (apart from interview non-respondents), but would skew findings if incorporated relative to the Stage 1 results. The analysis uses follow-up survey evidence from both stages in considering the potential magnitude of benefits and costs not measured monetarily.

[83]   The analysis eschews making any assumptions about (1) compensating changes to other types of government outlays, beyond the BOND interventions, in reaction to the fiscal position changes caused by the BOND

distortions by affecting incentives to work and invest. As explained in Appendix G, we draw from studies by economists of the size of these distortions per dollar of government taxing/spending to estimate the deadweight loss caused by the offset. Estimation of deadweight loss involves multiplying an estimate of the marginal excess tax burden (METB) times the estimated change in the government's fiscal position. Boardman et al. (Table 3.2, 2018) report a number of estimates of the METB. For the U.S., these range from 6 cents per dollar to 43 cents per dollar, with a median value of 19 cents per dollar.[84] It is thus 19 percent of the change in the government's fiscal position (as measured by other benefit and cost components in the analysis) that we use as the deadweight loss figure in the base case scenario.

It is important to note that, within the benefit-cost accounting framework used in this chapter, some benefits and costs do not affect society as a whole. For example, an increase in SSDI payments to beneficiaries causes a dollar-for-dollar offsetting loss for the Trust Fund. However, the cost of administering this transfer between beneficiaries and SSA is a cost that is not balanced by a benefit, and a worsening in the government's fiscal position incurs additional deadweight loss that results in a cost to society. Similarly, an increase in tax payments causes costs for beneficiaries that are exactly offset by the benefit to the units of government that receive those taxes—but again with deadweight loss implications.

6.1.3.   Accounting for the Time Dimension

Because a permanent offset policy would affect beneficiaries beyond the five years of the demonstration, the analysis projects benefits and costs for five years beyond the end of the evaluation data in 2015; i.e., the analysis imputes estimates for the years 2016 to 2020. The resulting benefit-cost study period of 10 years corresponds to the period used by the Congressional Budget Office to score proposals to Congress. Our analysis is limited to 10 years because of the difficulty of projecting from data produced by the demonstration benefits and costs reliably for SSDI beneficiary cohorts as they age.[85]

---

interventions or (2) changes in long-run deficit levels. Such scenarios mix in policy actions in realms that the BOND evaluation cannot and does not address and would fundamentally alter the meaning of the benefit-cost findings. The neutral position to adopt is that smaller or larger government outlays due to the demonstration's intervention equate to smaller or larger tax collections—and hence deadweight gain or loss.

[84]   The studies that produce these figures use estimates of uncompensated labor supply elasticity and rely on general equilibrium models. We rely on uncompensated, rather than compensated, labor supply elasticity because most taxpayers who would pay higher taxes under a national benefit offset are unlikely to be compensated. The researchers who conducted the studies were particularly interested in distortions caused by the income tax system. Therefore, we assume that payroll taxes result in distortions that are like those caused by income taxes.

[85]   The projections cannot take account of the fact that members of the research sample who were between age 56 and 59 at random assignment will age out of eligibility for SSDI during the projection period. Once they leave SSDI, the offset is unlikely to have much impact on them. Because the projections are based on impacts that occurred during the observation period, this causes benefits and costs during the projection period to be overstated, but perhaps not by very much. During the projection period, some sample members will become ineligible for SSDI as they reach their full retirement age. Although those ages 55 to 59 at baseline account for 29 percent of the sample, they were all under age 66 (the age at which they no longer qualify for SSDI benefits) throughout the BOND evaluation period and will be under 66 for at least part of the projection period. Moreover, even while they received SSDI, BOND's impacts on them appear slight. For example, the estimated impact on SSDI benefits due during the observation period was $359 for beneficiaries 50 and older at random

Like the BOND evaluation impact estimates themselves, our analysis ignores induced entry. Considering individuals who more recently became SSDI beneficiaries would involve even more extensive imputation outside the historical-cohort-based structure of the demonstration and evaluation. For impacts on earnings, SSDI and SSI payments, and income and payroll taxes, for which we have annual estimates, we assume impacts occurred in each year of the projection period equal to the average measured impact in the final two years of the observation period (2014-2015). The final two years were used to allow for the possibility that after the benefit offset was introduced several years of adjustment were necessary before a steady state situation was reached.[86]

Impacts on costs incurred in administering the offset, supplying work incentives counseling, providing SSA's Ticket to Work Program, and providing vocational rehabilitation services are all measured over the entire observation period, rather than year-by-year. For these outcomes, the benefit cost analysis assumes that the impact was constant throughout the observation and projection periods.

To further reflect the aspect of time, we adjusted all the monetary estimates used in the analyses to 2016 dollar equivalents using the *Consumer Price Index for Urban Wage Earners and Clerical Workers* (SSA n.d., accessed November 16, 2017). In addition, all benefits and costs have been discounted to a base year of 2016 using SSA's current 2.7 percent real discount rate.[87] This adjustment is necessary to reflect the fact that benefits received later are of less value than those received earlier and that costs incurred later are of less concern than those incurred earlier.

6.1.4.   Limitations of the Analysis

Several limitations apply to the benefit-cost analyses presented in this chapter. Most—but not all—can be addressed through sensitivity analyses, checking whether the thrust of the benefit-cost findings changes when uncertainties or assumptions are handled differently than in the "base case" scenario. Previous discussion has noted that some benefits and costs cannot be monetized, and that the implications of this fact will be assessed. To consider the consequences to the findings of the somewhat uncertain extrapolation of interventions' impacts for five years beyond the observation period, we compare the 10-year findings to results based exclusively on the five-year observation period.

Another factor built into the base scenario whose implications are explored is the decision to treat dollars as equivalent, regardless of who receives or pays them. On average, incomes of SSDI beneficiaries are below those of the taxpayers who incur the costs of the benefit offset policy; according to Wright et al. (2012) in the 2010 National Beneficiary Survey, around half of SSDI beneficiaries have incomes below the federal poverty line. As a result, it is likely that beneficiaries, on average, place greater value on an

---

assignment and $949 for those under 50. Thus, BOND's impact on those between 56 and 60 at random assignment will likely gradually fall during the projection period, but from an already low base.

[86]   The estimated impacts on earnings and SSDI payments shown in Exhibits G-2 and G-5, which are larger during the final two years of the observation period than during the first three years, suggests the presence of an adjustment period.

[87]   This is not the nominal discount rate, which includes the rate of inflation added to the average real rate of return on investments in the economy. The effect of inflation on measured dollar amounts in this benefit-cost analysis is removed by indexing to the 2016 price level as just described. The Social Security Board of Trustees assume a real discount rate of 2.7 percent for their intermediate-cost projections (SSA, 2018, Table II.C1).

additional dollar of income and consumption than do non-beneficiary taxpayers. A considerable literature exists suggesting that this difference should be dealt with in benefit-cost analysis by giving each dollar of gain or loss for individuals with relatively low incomes greater weight than each dollar of gain or loss for persons with higher incomes (see Boardman et al. 2018). To address this issue, we look at whether the findings are sensitive to using such a weighting scheme.

In addition, we examine whether the results are sensitive to the assumptions needed to estimate some of the individual benefits and costs—for example, the numeric factor drawn from the literature to compute deadweight loss as a proportion of the change in the government's fiscal position—by making alternative plausible assumptions.

As with the impact findings, most of the benefit and cost measures in the current chapter derive from data on individual study subjects. Thus, they contain sampling variability arising from chance factors in the selection of the demonstration sites and the assignment of individuals to treatment and control groups. Different sampling draws would have given different numbers. To take this uncertainty into account in each benefit-costs analysis conducted, we conduct a Monte Carlo analysis that—by using random draws in place of each initial benefit and cost amount[88]—produces 2,000 versions of the findings. Net benefits from each of the four perspectives are then averaged across the 2,000 scenarios and the fraction of the results that is positive is used to estimate the probability that the BOND interventions resulted in a net gain.

As discussed in Chapter 4, the benefit offset affected overpayments of SSDI benefits and, by changing the extent of overpayments, may also have altered the dollar amount of overpayments recovered by SSA. We have data to measure the first of these quantities but not the second. It is the net of these two components that matters to beneficiary and Trust Fund benefits and costs. Lacking the ability to measure that overall effect, we instead entirely exclude overpayments from the benefit-cost analysis by focusing on impacts on SSDI benefits due (rather than benefits paid, which would include overpayments but not their recovery).

The intent of the benefit-cost analyses is to assess the likely effects of a benefit offset operating in steady state. Because of implementation issues, the first few years of the demonstration period may not represent what occurred once operation of the benefit offset reached a steady state. We address this by performing a sensitivity analysis using a simulated scenario in which the average annual benefits and costs observed during the last two years of the observation period are assumed to have occurred in each of the first three years as well. By this means, findings that are more plausibly steady-state throughout the entire window of analysis can be compared to the main findings to estimate the difference start-up factors make to the research conclusions. In addition, the analysis excludes one-time-only costs incurred when rolling out BOND early in the demonstration—for example, resources used for mailings to Stage 1 treatment group subjects and for modifications to computer programs that determine SSDI benefits.

Finally, because the benefit-cost analyses, like the impact analyses, are based on a cohort of individuals who received SSDI benefits at the start of the demonstration in 2011, it does not take account of entry onto the SSDI rolls in later years. A benefit offset as a permanent policy could have an effect in that

---

[88]   The 2,000 random draws of each benefit and cost take the covariances among the various benefit and cost estimates fully into account.

period, inducing persons who were not on the SSDI rolls initially to enter in greater numbers. Because new entrants after 2011 were not included in the BOND research samples, induced entry is not considered in the evaluation overall, nor in the benefit-cost analysis in particular.

6.2. Stage 1 Base-Case Benefit-Cost Findings

Exhibit 6-1 presents 10-year "base-case" findings from the Stage 1 benefit-cost analysis—that is, it reports findings based on the set of assumptions that we judge to be the most plausible and, hence, produces what arguably might be called the "best estimates." Section 6.3 below examines the robustness of these base-case findings to variations in those assumptions. As discussed above, the 10-year findings incorporate both the five-year observation period and a five-year projection period. As previously indicated, the estimates are:

- Reported on a per beneficiary basis,

- Discounted to 2016 using a 2.7 real discount rate, and

- Adjusted to 2016 prices.

Benefits and costs from four perspectives appear in separate columns of the exhibit: SSDI beneficiaries, the Disability Insurance Trust Fund, other government, and all of society. Those benefits and costs that could be monetized appear in the top panel of the exhibit and are summed at the bottom of that panel. The bottom panel lists six benefits and costs that could be potentially important, but that could not be monetized for the Stage 1 analysis, and provides a judgement as to their likely size.

6.2.1. Benefits and Cost per Beneficiary

Estimates from the benefit cost analysis indicate that the Stage 1 intervention of BOND produced a per-beneficiary net benefit of $1,578 over 10 years, and a net cost of $1,589 for the Disability Insurance Trust Fund (see bolded center row of Exhibit 6-1). The impact of the benefit offset policy on SSDI benefits drives both results. Mostly because of taxes on increased SSDI payments, government entities other than the Disability Insurance Trust Fund are made slightly better off. Taking account of the deadweight loss to society (see below), society as a whole experienced a 10-year net cost of $188 per beneficiary from providing the benefit offset to the full SSDI caseload.

It is worth noting that the individual monetized benefits and costs that lead to these totals are very small for figures encompassing 10 years, in substantial part because less than four percent of beneficiaries used the benefit offset during the first five years. Moreover, because SSDI benefits and taxes on SSDI benefits have offsetting effects on beneficiaries and government, they do not show up in the column for society as a whole. However, the increase in taxes needed to support the increase in SSDI payments causes a small amount of deadweight loss that appears in the column for society. While the deadweight loss is small ($282, or about $28 per beneficiary per year over 10 years), it is more than enough to outweigh what would otherwise be an even smaller positive gain from society's perspective.

Exhibit 6-1.   Benefits and Costs of the Stage 1 Benefit Offset, by Accounting Perspective Over 10 Years in 2016 Present Value

Benefit or Cost Component	Beneficiaries	Disability Insurance Trust Fund	Other Government	All of Society
Monetized				
Pre-tax earnings	$73	$0	$0	$73
Fringe benefits from work	34	0	0	34
SSDI benefits	1,584	-1,584	0	0
SSI benefits	-12	0	12	0
SSI administrative costs	0	0	1	1
Payroll taxes	-10	10	0	0
Income taxes	-20	0	20	0
Sales taxes	-38	0	38	0
SSDI/BOND administrative costs	0	-1	0	-1
Cost of Ticket-to-Work	0	-14	0	-14
State VR service costs (net of Ticket-to-Work)	0	0	35	35
Work-related expenses (e.g., child care, transportation, clothing)	-7	0	0	-7
Non-market time	-27	0	0	-27
Deadweight loss	0	0	0	-282
Net Monetized Benefits (+) / Costs (-)	**1,578**	**-1,589**	**106**	**-188**
Non-Monetized—with Likely Magnitude				
Value placed by public on increasing work among SSDI beneficiaries	0	0	0	Probably negligible
Labor market effects on third parties	0	0	0	Probably negligible
Value of output from voluntary work	0	0	0	Negative, but probably small
Health status and life-satisfaction	Probably negligible	0	0	Probably negligible
Earnings of spouses	Probably altered less than SSDI benefits	0	0	Probably altered less than SSDI benefits
Other government benefits	Probably small, direction uncertain	0	Probably small, direction uncertain	0

*Rows and columns may not sum to totals due to rounding.

Notes: See Section 6.1 and Appendix G for discussion of each benefit/cost component. All benefits and costs are dollars per beneficiary over ten years and are inflation-adjusted to 2016 dollars and discounted to 2016 present value.

Unweighted sample sizes: T1 = 77,101; C1 = 891,429.

Other government costs from the offset policy—e.g., increases in the cost of operating state vocational rehabilitation programs—are quite small. The costs of administering the offset were only trivially more than the costs of administering current law (less than a dollar over 10 years), an estimate that is not statistically significant.[89] Our estimates suggest that the offset policy causes operating costs of the Ticket-to-Work program to fall slightly.

In sum, when applied to the full range of SSDI beneficiaries in Stage 1, the benefit offset might be best viewed as a very modest transfer program, one that increases the incomes of some employed beneficiaries at the expense of other workers who contribute to the Trust Fund and presumably have somewhat higher incomes. Society obtains this redistribution at a deadweight cost of 18 percent (i.e., $1 in deadweight loss for every $5.60 of redistribution achieved).[90]

Would the conclusions based on the top panel Exhibit 6-1 have changed had the monetary values of the items listed in the bottom panel been estimated so that they could have been included in the monetary calculus?[91] Probably not. The value the public places on increasing work among SSDI beneficiaries and on labor market effects on third parties (for example, a reduction in jobs available to non-beneficiaries because they were filled by beneficiaries) depends on BOND having increased employment among SSDI recipients. Measured impacts on earnings-related outcomes are modest for Stage 1, occurring in just occasional years and for only some earnings ranges relative to BYA. Measured impacts never involve annual employment rate increases or decreases of more than 0.3 of a percentage point (Exhibit F-1). Hence, potential effects on these benefit-cost items are listed in the exhibit as "probably negligible." This is admittedly conjectural, however. We simply do not have the information to draw firm conclusions.

A decline in voluntary work might have occurred had the BOND benefit offset substantially increased hours of paid work. Impact findings for any paid work and for earnings show such small effects that an important amount of displacement of volunteer work seems unlikely.[92] The increase in beneficiary incomes (mostly from higher SSDI payments) would be expected to have had a positive effect on health

---

[89]   Stage 1 impact estimates for distinct components of administrative costs are shown in Appendix G, Exhibit G-4.

[90]   The $5.60 figure is calculated by dividing the net benefit to beneficiaries ($1,578) by the deadweight loss ($282).

[91]   The 36-month follow-up survey was not suitable for estimating specific benefits and costs for Stage 1, because the sample size (approximately 3,000 respondents in each of T1 and C1) was not large enough to produce meaningful estimates of impacts that were presumably quite small. The full Stage 1 samples are much larger— about 77,000 for T1 and 900,000 for C1—because very large samples are necessary to detect small but substantively important impacts when the intervention affects the behavior of only a small share of the studied subjects. For these reasons, the analysis cannot consider monetary values of Stage 1 impacts on the following outcomes: spousal earnings, time spent in voluntary work, and receipt of government transfer payments other than SSDI and SSI (for example, SNAP, TANF, veterans' benefits, disability insurance for a disabled adult child, and unemployment insurance). Thus, the top panel in Exhibit 6-1 does not include the monetary values of BOND-induced changes in these outcomes. However, they are listed in the bottom panel of the exhibit and discussed here.

[92]   The estimated impact on hours of volunteer time per week was -0.2 in the 36-month follow-up survey for Stage 1, and this estimate was not statistically significant

---

and life satisfaction,[93] but the SSDI benefit increase was so small, well under $200 per year over 10 years, that it is difficult to imagine much of an effect here.[94] Conceivably the offset could have affected beneficiary mortality, though the data show no indication of an effect on the death rate (the T1 and C1 groups differ by just 0.02 of a percentage point for this outcome, which is only 0.2 percent of the control group rate; see Appendix F, Exhibit F-16). With so little to suggest any health or mortality effect, we list health status and life-satisfaction in Exhibit 6-1 as "probably negligible" effect.[95]

Another possible area of costs and benefits concerns the spouses of beneficiaries. It is possible that beneficiaries' marriage partners worked less or more because the earnings and SSDI benefits of their husbands or wives changed. However, not only was the increase in beneficiary incomes from these sources small—around $170 per year, mostly from larger SSDI benefits—less than a third of SSDI beneficiaries are married (Wright et al. 2012). Certainly, one would not expect a spousal earnings response greater on average than the average change in SSDI benefits, given the large share of unmarried beneficiaries for whom no spousal earnings response was possible.

Due to the benefit offset policy, government benefit payments could potentially have changed in a variety of programs in addition to SSDI and SSI. Had the offset appreciably increased earnings and employment, for example, receipt of unemployment insurance benefit payments might have increased, while payments under TANF and SNAP (Food Stamps) might have decreased. However, as previously emphasized, BOND had little effect on employment and earnings needed to produce unemployment insurance payment changes. Benefits from transfer programs such as TANF and SNAP can be expected to have declined due to the increase in SSDI but by well under a dollar for every dollar by which SSDI increased.[96] Moreover, many of those who received increases in SSDI benefits probably were not on TANF and SNAP. Beneficiaries receiving increases in SSDI benefits had to have positive earnings, making them a subset of subjects less likely to be enrolled in TANF or SNAP.[97] Further, the threat of reductions in TANF or SNAP benefits presumably would discourage treatment subjects enrolled in those

---

[93]   Both increased work and increased income tend to improve people's health and life-satisfaction (Fujiwara 2010). Evans and Moore (2011), however, find that mortality increases with income in the short-run because increased income results in increased consumption.

[94]   There could have been substantial effects on the health and life-satisfaction of some of those who actually used the benefit offset, but because users comprise less than 4 percent of the sample when averaged over the entire sample, the overall effects are likely to be very small. The 36-month follow-up survey for Stage 1 included several measures of health. The estimated impacts on these measures vary in sign and are generally of negligible size and statistically insignificant. There was no detectable impact on self-rated overall health and self-reported measures of physical functioning, emotional health, and Body Mass Index.

[95]   See additional analysis of mortality in Appendix I.

[96]   Based on the 36-month survey, the estimated impacts on the Stage 1 treatment group's receipt of payments under various government transfer programs are small (in all instances under $50 a year), of varying sign, and in no case statistically significant. This is not surprising because, except for SNAP, few SSDI beneficiaries participate in these programs. While a third of the Stage 1 treatment group members received SNAP benefits at the time of the survey, the estimated impact on benefit payments was small (a decrease of around $3 a year) and statistically insignificant.

[97]   One-third (33 percent) of T1 beneficiaries indicated that they received SNAP benefits at the time of the 36-month survey. Only 3 percent received TANF.

---

programs from taking advantage of the benefit offset by increasing their earnings—making offset-induced reductions in TANF and SNAP benefits even less likely. In total, we judge the potential impact of the BOND intervention on other government benefits to be small and of uncertain direction.

6.2.2.   Projection of Net Benefits to the Nation

While the per beneficiary benefit and cost estimates in Exhibit 6-1—which pertain to the average SSDI beneficiary in the U.S. on the rolls in 2011—are quite small, the number of SSDI beneficiaries is very large. This implies that aggregate benefits and costs from future use of a benefit offset policy could be substantial. For example, the Stage 1 research sample represents the full population of 6,453,341 SSDI beneficiaries who were ages 21 to 59 in 2011. Multiplying the per-beneficiary estimates (from Exhibit 6-1) by this total number of beneficiaries yields the aggregate net benefit that would have been produced (in 2016 dollars) by applying the offset to all those beneficiaries. This extrapolation implies a total benefit to SSDI beneficiaries of $10.2 billion over the 10 years from 2011 to 2020 (Exhibit 6-2). Over the same period, however, a nationwide intervention like BOND is estimated to draw down the Disability Insurance Trust Fund by $10.3 billion, an amount almost entirely attributable to the increase in SSDI payments.[98] Due to a deadweight loss of 1.8 billion, society would sustain a net loss estimated at $1.2 billion even though other parts of the government would gain $0.7 billion.

Exhibit 6-2.   Stage 1 Net Benefits by Scale and Accounting Perspective over 10 Years

	SSDI Beneficiaries	Disability Insurance Trust Fund	Other Government	All Society
Per Beneficiary	$1,578	-$1,589	$106	-$188
Aggregated across all beneficiaries[a]	$10.2 billion	-$10.3 billion	$0.7 billion	-$1.2 billion

Notes: All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W), and discounted to 2016 present value using a real discount rate of 2.7 percent.

Unweighted sample sizes: T1 = 77,101; C1 = 891,429.

[a] Beneficiaries represented by the Stage 1 sample include 6,453,341 individuals ages 21 to 59 in 2011. The estimates in the second row were computed as the products of 6,453,341 and the estimates in the first row.

6.3.   Robustness of the Base-Case Findings for Stage 1

As previously discussed, various limitations apply to the base-case benefit-cost results that appear in Exhibit 6-1. To explore the importance of these limitations for the substantive conclusions above, we performed several sensitivity analyses. This section presents the results of those analyses.

---

[98]   The actual aggregate figures would be even larger if the benefit offset were made available to beneficiaries between the ages of 60 and 66. In 2015, there were 10,237,204 total SSDI beneficiaries with the difference between this number and the 6,453,341 beneficiaries who were ages 20 to 59 in 2011 mainly due to persons between the ages of 60 to 65. Assuming that the benefit-cost findings are applicable to these persons as well as to those who are younger, the 10-year cost of an offset to the Trust Fund would be $16.3 billion, a figure that slightly exceeds the $16.2 billion in gains estimated for beneficiaries. A further gain of $1.1 billion would accrue to various units of government. However, because of a $2.9 billion deadweight loss society comes out behind by an estimated $1.9 billion over 10 years.

6.3.1.   Sensitivity to Projection Assumptions

The analysis of 10-year benefits and costs in the base case assumes that impacts observed during the first five years of the demonstration would have continued at the same level for five more years were the offset kept in place. While this may be the best-guess projection, scrutiny of a less bold assumption concerning effects in the second five-year period shows the sensitivity of the benefit-cost analysis conclusions to the projections made for the years 2016 to 2020. The least bold assumption posits that no further benefits or costs of the offset arise after the end of the original 5-year observation period—i.e., that the offset would have zero effects in all areas in 2016 through 2020. Estimates under this assumption appear in Appendix G, Exhibit G-3. Like the base case results, these results again show gains for beneficiaries and other government units and losses for the Trust Fund, with society as a whole nearly breaking even. However, as would be anticipated, the dollar amounts of these net benefits and costs are considerably smaller than in the base case. Net monetized benefits from each perspective appear in Exhibit 6-3 under the two different projection scenarios.

Exhibit 6-3.   Stage 1 Net Benefits, by Perspective, for Alternative Projection Assumptions

Projection Assumptions	Beneficiaries	DI Trust Fund	Other Government	All of Society
5-year findings	$721	-$745	$42	-$116
10-year findings (steadily continued effects in years 6-10)	$1,578	-$1,589	$106	-$188

Notes: All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W), and discounted to 2016 present value using a real discount rate of 2.7 percent.

Unweighted sample sizes: T1 = 77,101; C1 = 891,429.

The same segments of society gain and lose in the two scenarios, but by smaller amounts in the scenario that assumes zero effects in years 6 to 10. The downward shift from the base-case reduces net benefits by a little more than half to beneficiaries and to other government, and reduces net costs by a little more than half to the Trust Fund. An argument can be made that the assumption in the base case scenario of annual effects from the observation period continuing another 5 years constitutes a middle-ground analysis between:

(a) No effects past the observation period (i.e., ignoring effects that likely would occur in years 6 through 10), and

(b) Extrapolating past ten years (say to 15 years).

Lacking the ability to directly measure benefits and costs into a sixth year and beyond, some projection assumption must be made; the assumption of just 5 more years at steady state seems reasonable.

6.3.2.   Distributional Weighting

A standard assumption in economics is that each additional dollar an individual receives provides less utility than the preceding dollar. Because SSDI beneficiaries on average have much lower incomes than the typical taxpayer, this assumption implies that beneficiaries should value a given dollar change in income more highly than the equivalent income change for the average taxpayer. We address this issue through a sensitivity analysis that applies a greater weight to the dollar gains received by beneficiaries

than to the losses incurred by taxpayers. The challenge in doing this is in finding an appropriate set of weights.

Fujiwara (2010) compared the findings from empirical studies that provided evidence about weights that might be used and concluded that almost all implied that the weight for relatively low-income persons is greater than 1 if the weight for high income individuals is set at 1.[99] An empirical study by Layard, Nickell, and Mayraz (2008) supports setting the weight for persons in the U.S. in the lowest two income quintiles at around 3, given a weight of 1 for higher income persons. A recent review of the literature by Boardman et al. (2018) concludes that a reasonable upper bound value for a relative distributional weight for the disadvantaged ranges from 1.5 to 2. A weight as small as 1.2 turns the estimated net societal benefits from Stage 1 of BOND from negative (-$188) to positive ($128). The benefits would become even more positive if the larger weights suggested by the literature were used. For example, a weight of two, which is at the upper end of the range suggested above, implies that net social benefits are $1,390 per beneficiary. Still, if this figure is divided by 10 to put it roughly on a per-year basis, it comes to less than $140 per year. Comparing this figure to the annual loss $19 (i.e., $188/10) if distributional weighting is ignored suggests that that society neither gains nor loses appreciably from the BOND benefit offset.

6.3.3.   Sensitivity to Additional Assumptions

Estimating the BOND interventions' benefits and costs requires numerous assumptions, described earlier in the text and in Appendix G. This results in uncertainty concerning the dollar values of certain benefits and costs. However, with one exception, the dollar values involved for the uncertain estimates are so small that even halving or doubling them would not affect the conclusions from the analysis[100] about net benefits or costs to various segments of society.[101]

---

[99]   Obtaining values for the weights requires an estimate of the elasticity of the marginal utility of income. One method that has been used to estimate this quantity relies on surveys in which respondents are asked their incomes and to rate their level of happiness on a scale (e.g., very happy, pretty happy, not too happy) then uses their responses as a proxy for utility.

[100] The one impact large enough to substantially alter the benefit-cost findings if doubled or halved—the offset's impact on SSDI benefits—is based directly on administrative data for the SSDI program and does not require special assumptions. Other than the sampling variability reflected in the standard error of the estimate of impact on SSDI benefits (which the extremely large Stage 1 sample size makes quite small), there is little uncertainty concerning it. Moreover, we address sampling variability in the Monte Carlo analysis reported below.

[101] For beneficiaries, the most important variations of this sort to consider are ones that might undercut the conclusion of substantial net benefits. Looking at variations that create the extreme worst-case scenario—a scenario that that halves all the positive estimates for beneficiaries in Exhibit 6.1, except for the estimate for SSDI benefits (whose robustness is discussed in the previous footnote and also below) and that doubles the negative estimates—we find that net benefits for beneficiaries decline from $1,578 to $1,411. Similarly, to determine whether the negative net benefit finding for the Disability Insurance Trust Fund holds up to variations in component benefits and costs, consider the best-case scenario that doubles the positive estimates in that column and halves the negative estimates, except for the estimate for SSDI benefits. Here, the bottom line for the Trust Fund scarcely changes, falling from $1,589 to $1,572. The parallel best-case exercise for entries in the society as a whole column, holding constant the estimate for deadweight loss (whose implications for the social bottom line are discussed below, for differing estimation approaches and values), reduces estimated net social cost from -$188 to -$21.

The exception is the estimate of deadweight loss, which was obtained by multiplying an estimate of the marginal excess tax burden (METB) by the estimated change in the government's fiscal position (as indicated by the total net benefit estimates in the columns in Exhibit 6-1 for the Disability Insurance Trust Fund and for other government entities). As noted earlier, for the U.S. estimates of the METB range from 6 cents per dollar to 43 cents per dollar, with a median value of 19 cents per dollar. The estimate of deadweight loss in the base case in Exhibit 6-1 uses the median value of 19 cents per dollar. Were 6 cents per-dollar substituted instead, the lowest estimate reported by Boardman et al. (2018), deadweight loss would shrink from -$282 to -$89 and estimated societal net benefits would be very close to zero, changing from -$188 to +$5. In contrast, if the METB was 43 cents per dollar (the highest estimate reported by Boardman et al.), costs resulting from deadweight loss would grow to $638 and the cost of the BOND benefit offset to society would increase to $544. Thus, the estimate of the cost to society is somewhat sensitive to the assumed value of the METB used to compute deadweight loss. However, the scenarios that result in almost no loss to society require very small METB values at the lower extreme of those found in the literature.

Earlier chapters suggest that it took time to fully implement the benefit offset. In fact, Exhibit G-2 indicates that the benefit offset's impacts on SSDI payments increased over the observation period. Presumably, this would not occur in an on-going program.[102] To get a sense of how start-up implementation may have affected the benefit-cost findings, we examine in Exhibit 6-4 what would have happened if program impacts in 2011, 2012, and 2013 had equaled the average impact for two years close to full implementation and steady state, 2014 and 2015—holding the projections for 2016 through 2020 constant. As shown in the exhibit, this "steady-state" scenario makes benefits to beneficiaries and costs to the Trust Fund larger, but not by very much (by 16 and 14 percent, respectively). Because net benefits to beneficiaries rise by a little more than the increase in costs to the Trust Fund, societal net benefits become slightly less negative. These results suggest that the base case is a reasonable approximation of steady state, without making special assumptions about alternative benefit and cost amounts in the earlier start-up years of the demonstration.

Exhibit 6-4.   **Stage 1 Net Benefits, by Perspective, Modeling Steady State**

Model	Beneficiaries	DI Trust Fund	Other Government	All of Society
Base-case (actual results in early years)	$1,578	-$1,589	$106	-$188
No implementation variation	$1,838	-$1,811	$136	-$155

Notes: The no implementation variation scenario imputes average impacts over 2014 and 2015 for the first three years of the demonstration. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W), and discounted to 2016 present value using a real discount rate of 2.7 percent.

Unweighted sample sizes: T1 = 77,101; C1 = 891,429.

The various benefits from and costs of BOND accrue at different points in time. In as much as people value a dollar next year differently than a dollar ten years from now, it is inappropriate to simply sum

---

[102]   Although the growth in the size of the impacts on SSDI benefits over time suggests that it took time for BOND to reach a steady state, the growth could also reflect the aging of the research sample (if that mattered to impacts) or changes over time in external economic conditions.

dollars across years. Instead, we apply a real discount rate to compute the present value in 2016, after which they are comparable and can be summed. (Before this is done, price differences between years—due to inflation—are removed by denominating everything in 2016 dollars.) Our base-case estimates use a real discount rate of 2.7 percent, the rate SSA actuaries currently use to convert 2011-2015 and 2017-2020 dollar amounts into their 2016 value equivalents (SSA 2018). However, the proper real discount rate to use in benefit-cost analysis is controversial (Boardman et al. 2018, Chapter 10). In a detailed review of the literature, Boardman et al. (2018, Chapter 10) suggests a real discount rate of 3.5 percent—close to SSA's 2.7 percent rate—and recommends conducting sensitivity analysis with rates of 2.3 and 5.5 percent. Exhibit 6-5 summarizes the implications of using these alternative rates for the base-case findings.

Exhibit 6-5.   Stage 1 Net Benefits, by Perspective, with Different Real Discount Rates

Real Discount Rate	Beneficiaries	DI Trust Fund	Other Government	All of Society
2.3%	$1,577	-$1,588	$106	-$187
2.7% (base-case)	$1,578	-$1,589	$106	-$188
3.5%	$1,580	-$1,592	$105	-$190
5.5%	$1,587	-$1,602	$104	-$196

Notes: All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W), and discounted to 2016 present value using various real discount rates (as indicated in the exhibit).
Unweighted sample sizes: T1 = 77,101; C1 = 891,429.

As can be seen, the choice of real discount rate makes almost no difference to the findings, even when the rate is more than doubled. One reason for this is that the 10-year period over which the discounting takes place is relatively short. The short time horizon limits the year-by-year compounding of the discounting process. Also, even the highest rate creates only modest compounding in the time span covered (at most five years of adjustment forward or backward in time to the 2016 accounting year used in the analysis).[103] Further, discounting all quantities to the middle of the 10-year analysis period has virtually offsetting effects on benefit and cost measures that remain fairly stable from year to year.[104]

6.3.4.   Accounting for Sampling Variation Through Monte Carlo Analysis

As noted earlier, benefit and cost figures derived from data on individual study subjects are subject to sampling variability arising from chance factors in the selection of the demonstration sites and the assignment of individuals to treatment and control groups. This means that different sampling draws

---

[103] Five years of discounting, from 2020 back to 2016, gives a discount factor of 0.81 at a compounded 5.5 percent annual real discount rate. (A 2016 amount of 81 cents accumulates to 1 dollar by 2020 with an annual rate of return of 5.5 percent, going from 81 cents to 85 to 90 to 95 to 100.) Moving in the other direction, a 5.5 annual rate implies that 1 dollar received in 2011 rises in value to $1.31 by 2016, meaning that a multiplier of 1.31 gets applied. The full range of discount/multiplier factors—0.81 to 1.31—is not large as such things go and, with real discount rates of 3.5 or less, becomes much smaller.

[104] Measures that apply to years after 2016 are discounted by a factor smaller than one; while those from years prior to 2016 have discount factors that are larger than one. Applying these two sets of discount factors to a reasonably steady dollar amount for all years creates offsetting effects.

would have produced different numbers. To take this uncertainty into account, we conduct a Monte Carlo analysis that—by using random draws in place of each initial benefit and cost measure[105]—produces 2,000 versions of the Stage 1 benefit-cost findings. Net benefits from each of the four perspectives are then averaged across the 2,000 cases and the fraction of the results that are positive is used to estimate of the probability that the intervention produced net gains from each perspective. Exhibit 6-6 provides the findings from this analysis.

Exhibit 6-6.   Sensitivity of Stage 1 Estimates to Sampling Variability

	Perspective			
	Beneficiaries	**DI Trust Fund**	**Other Government**	**All of Society**
Base-case net benefits	$1,578	-$1,589	$106	-$188
Average net benefits from 2,000 sampling draws	$1,581	-$1,589	$106	-$185
Standard deviation of net benefits from 2000 sampling draws	$221	$193	$81	$170
95-percent confidence interval for net benefits	$1,145 to $2,011	-$1,968 to -$1,211	-$54 to $265	-$522 to $146

Notes: The 95-percent confidence intervals are centered on the base-case net benefits. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W), and discounted to 2016 present value using a real discount rate of 2.7 percent.

Unweighted sample sizes: T1 = 77,101; C1 = 891,429.

Average net benefits from the four different perspectives, derived by averaging values across 2,000 Monte Carlo sampling draws, are very similar to the original base-case estimates. This is unsurprising because each Monte Carlo draw represents a random deviation from the original measures.

The important new information in the exhibit appears in the third row—standard deviations that characterize the extent to which net benefit conclusions depend on the particular sampling draw that occurred in the demonstration.[106] The standard deviations shown for the beneficiary and the Disability Insurance Trust Fund perspectives are small relative to the associated average net benefit figures. This means that 95-percent confidence intervals for the net benefit amounts—defined as 1.96 standard deviations on either side of the average—do not include zero (see the bottom row of the exhibit). In fact, none of the 2,000 trials produced a negative net benefit estimate from the beneficiary perspective, nor a positive net benefit estimate from the perspective of the Trust Fund. This finding occurs because, as Exhibit 6-1 shows, net benefits from both perspectives almost entirely reflect the offset's estimated impact on SSDI benefits, which the data estimate with high statistical precision (see Exhibit G-2). The boundaries of the 95-percent confidence interval (see bottom row of Exhibit 6-6) provide a strong basis

---

[105]   The 2,000 random draws of each benefit and cost take the covariances among the various benefit and cost estimates fully into account.

[106]   Sixty-six percent of all possible draws will produce standard deviations that, when added to or subtracted from the base-case figure, contain the true net benefit amount.

for concluding[107] that—notwithstanding the potential consequences of sampling variability—the Stage 1 benefit offset generated positive net benefits for beneficiaries of over $100 per year ($1,145 or more over 10 years) and perhaps $200 per year (if the total gain in 10 years equals the $2,011 upper limit of the confidence interval). Similarly, one can have confidence that the intervention imposed a net cost on the Disability Insurance Trust Fund of at least $100 per year ($1,211 or more over 10 years) and perhaps almost $200 per year (given the $1,968 lower limit of the confidence interval).

There is considerably greater uncertainty concerning net benefits for other government entities and, more importantly, society as a whole. With large standard deviations relative to their averages, the 95-percent confidence intervals for other government and societal net benefits both include zero. Thus, as shown in comparing the top and bottom rows of Exhibit 6-6 for these perspectives, net benefits opposite in sign to the base case finding are plausible. Even so, other government net benefits are likely to be positive and those for society as a whole negative. For example, only 13 percent of the 2,000 sampling draws resulted in a positive net benefit estimate for society, implying that there is an 87 percent probability that the Stage 1 benefit offset imposed net costs on society. Importantly, however, those costs are unlikely to be substantial: they exceeded $300 per beneficiary in only 4 percent of the trials. Furthermore, when a distributional weight of 1.2 is used (see Section 6.3.3), societal net benefits are positive in 60 percent of the sampling draws; and with a weight of 1.5, societal net benefits exceed zero in 96 percent of the draws. The bottom row findings in Exhibit 6-6 show it to be even less likely that benefits to other government units of an appreciable magnitude took place.

6.4.   Stage 2 Base-Case Benefit-Cost Findings

Next, we consider benefit-cost findings from Stage 2 of BOND. Exhibit 6-7 presents 10-year "base-case" findings from the Stage 2 benefit-cost analysis for the offset-plus-WIC group (T21) compared to a control group (C2) that is subject to current law. Exhibit 6-8 presents 10-year "base-case" findings from the Stage 2 benefit-cost analysis for the offset-plus-EWIC group (T22) compared to a control group (C2) that is subject to current law. Exhibit 6-9 presents 10-year "base-case" findings from the Stage 2 benefit-cost analysis for the offset-plus-EWIC treatment group compared to the offset-plus-WIC treatment group (T22 compared to T21). These findings are derived using methods parallel in all respects to the methods underlying the earlier Stage 1 results.

As for Stage 1, the overarching finding for Stage 2 is that the offset policy results in gains for beneficiaries and losses for the Disability Insurance Trust Fund. For the offset-plus-WIC group (T21) compared to a control group (C2), there are net gains for other government units (aside from the Disability Insurance Trust Fund) due to increased tax collections. However, for the offset-plus-EWIC group (T22) compared to a control group (C2), there are no net gains (in fact, a tiny net loss) for other government units because increased tax collections are insufficient to offset increased spending on vocational rehabilitation services.

The per-beneficiary gains to society over 10 years, relative to current law, are estimated at almost $3,000 for the offset plus WIC (Exhibit 6-7). But the per-beneficiary losses to society over 10 years, relative to

---

[107]   As noted earlier, findings from the Bayesian statistical framework used in this benefit-cost analysis can be stated in terms of likelihood—i.e. how strong one's beliefs can be about a particular benefit or cost being of a particular size, given an agnostic starting point and the sampling variability contained in one's data.

current law, are estimated at over $5,000 for the offset plus EWIC (Exhibit 6-8). In Exhibit 6-9, we see that the net loss to society over 10 years from implementing the offset plus EWIC instead of the offset plus WIC, would be in excess of $8,000 per beneficiary.

Relative to current law, the net benefits for SSDI beneficiaries are positive and substantial: a gain of $7,642 from the offset plus WIC compared to current law and $8,363 from the offset plus EWIC. Net losses to the Trust Fund due to Disability Insurance payments alone are in the $5,000 to $6,000 per beneficiary range in both treatment arms. Gains to beneficiaries are slightly larger with the more intensive EWIC services rather than standard WIC services. But the losses to the Trust Fund are much larger with the intensive EWIC services because of the costs associated with delivering counseling, so the offset plus EWIC results in net costs to society, in contrast to the net benefits from the offset plus WIC. The offset plus EWIC incurs an additional almost $6,000 in counseling costs per beneficiary over 10 years compared to the offset plus WIC. The larger societal losses for the offset plus EWIC are also due to increased costs for other government entities—particularly the state VR agencies.

These net benefits exclude benefits and costs for which we do not have monetary values. Would the conclusions in the top panels of Exhibits 6-7, 6-8, and 6-9 change had the monetary values of the items listed in the bottom panels been estimated so that they could have been included in the monetary calculus?[108] Probably not. The value the public places on increasing work among SSDI beneficiaries and on labor market effects on third parties (for example, a reduction in jobs available to non-beneficiaries because they were filled by beneficiaries) depends on the Stage 2 BOND interventions having increased employment among SSDI recipients. Measured impacts on employment are modest for the different Stage 2 policy comparisons, occurring in only some years and for some earnings ranges relative to BYA (Exhibit F-7). None of the measured effects on employment rates show an increase or decrease of more than 3 percentage points. Hence, potential effects on these benefit-cost items are listed in the exhibit as "probably negligible." This is admittedly conjectural. We simply do not have sufficient information to draw firm conclusions with regard to the non-monetized components.

---

[108]   The Stage 2 36-month follow-up survey was not used to estimate any of these benefits and costs for inclusion in the bottom-line monetary findings, consistent with the Stage 1 approach. However, as in Stage 1, information from the Stage 2 survey is invoked to supplement the discussion of non-monetary benefits and costs. Benefit-cost results partially based on the Stage 2 survey are presented and discussed in Appendix G. These results are qualitatively similar to those based solely on administrative data presented in this chapter.

Exhibit 6-7.   **Benefits and Costs of the Offset Plus WIC versus Current Law in Stage 2, by Accounting Perspective over 10 Years in 2016 Present Value**

Benefit or Cost Component	Beneficiaries	Disability Insurance Trust Fund	Other Government	All of Society
Monetized				
Pre-tax Earnings	$3,788	$0	$0	$3,788
Fringe benefits from work	1,750	0	0	1,750
SSDI benefits	5,087	-5,087	0	0
SSI benefits	22	0	-22	0
SSI administrative costs	0	0	-2	-2
Payroll taxes	-566	566	0	0
Income taxes	-529	0	529	0
Sales taxes	-184	0	184	0
SSDI/BOND administrative costs	0	5	0	5
Counseling costs	0	-43	0	-43
Cost of Ticket-to-Work	0	-239	0	-239
State VR service costs (net of Ticket-to-Work)	0	0	193	193
Work-related expenses	-341	0	0	-341
Non-market time	-1,384	0	0	-1,384
Deadweight loss	0	0	0	-744
Net Monetized Benefits (+) / Costs (-)	**7,642**	**-4,798**	**882**	**2,982**
Non-Monetized—with Likely Magnitude				
Value placed by public on increasing work among SSDI beneficiaries	0	0	0	Probably negligible
Labor market effects on third parties	0	0	0	Probably negligible
Value of output from voluntary work	0	0	0	Probably negligible
Health status and life-satisfaction	Probably negligible	0	0	Probably negligible
Earnings of spouses, other income	Assumed negligible	0	0	Assumed negligible
Other government benefits	Assumed negligible	0	Assumed negligible	0

Notes: See Section 6.1 and Appendix G for discussion of each benefit/cost component. All benefits and costs are dollars per beneficiary over ten years and are inflation-adjusted to 2016 dollars and discounted to 2016 present value.

Unweighted sample sizes: T21 = 4,854, T22 = 3,041, C2 = 4,849.

Exhibit 6-8.   Benefits and Costs of the Offset Plus EWIC versus Current Law in Stage 2, by Accounting Perspective over 10 Years in 2016 Present Value

Benefit or Cost Component	Beneficiaries	Disability Insurance Trust Fund	Other Government	All of Society
Monetized				
Pre-tax Earnings	$4,078	$0	$0	$4,078
Fringe benefits from work	1,884	0	0	1,884
SSDI benefits	5,490	-5,490	0	0
SSI benefits	23	0	-23	0
SSI administrative costs	0	0	-2	-2
Payroll taxes	-609	609	0	0
Income taxes	-444	0	444	0
Sales taxes	-201	0	201	0
SSDI/BOND administrative costs	0	-60	0	-60
Counseling costs	0	-5,914	0	-5,914
Cost of Ticket-to-Work	0	-491	0	-491
State VR service costs (net of Ticket-to-Work)	0	0	-648	-648
Work-related expenses	-367	0	0	-367
Non-market time	-1,490	0	0	-1,490
Deadweight loss	0	0	0	-2,161
Net Monetized Benefits (+) / Costs (-)	**8,363**	**-11,345**	**-27**	**-5,170**
Non-Monetized—with Likely Magnitude				
Value placed by public on increasing work among SSDI beneficiaries	0	0	0	Probably negligible
Labor market effects on third parties	0	0	0	Probably negligible
Value of output from voluntary work	0	0	0	Probably negligible
Health status and life-satisfaction	Probably negligible	0	0	Probably negligible
Earnings of spouses, other income	Assumed negligible	0	0	Assumed negligible
Other government benefits	Assumed negligible	0	Assumed negligible	0

Notes: See Section 6.1 and Appendix G for discussion of each benefit/cost component. All benefits and costs are dollars per beneficiary over ten years and are inflation-adjusted to 2016 dollars and discounted to 2016 present value.

Unweighted sample sizes: T21 = 4,854, T22 = 3,041, C2 = 4,849.

Exhibit 6-9. Benefits and Costs of EWIC versus WIC, Given the Offset, in Stage 2, by Accounting Perspective over 10 Years in 2016 Present Value

Benefit or Cost Component	Beneficiaries	Disability Insurance Trust Fund	Other Government	All of Society
Monetized				
Pre-tax Earnings	$290	$0	$0	$290
Fringe benefits from work	134	0	0	134
SSDI benefits	403	-403	0	0
SSI benefits	1	0	-1	0
SSI administrative costs	0	0	0	0
Payroll taxes	-43	43	0	0
Income taxes	85	0	-85	0
Sales taxes	-17	0	17	0
SSDI/BOND administrative costs	0	-65	0	-65
Counseling costs	0	-5,871	0	-5,871
Cost of Ticket-to-Work	0	-252	0	-252
State VR service costs (net of Ticket-to-Work)	0	0	-841	-841
Work-related expenses	-26	0	0	-26
Non-market time	-106	0	0	-106
Deadweight loss	0	0	0	-1,417
Net Monetized Benefits (+) / Costs (-)	**721**	**-6,547**	**-909**	**-8,152**
Non-Monetized—with Likely Magnitude				
Value placed by public on increasing work among SSDI beneficiaries	0	0	0	Probably negligible
Labor market effects on third parties	0	0	0	Probably negligible
Value of output from voluntary work	0	0	0	Probably negligible
Health status and life-satisfaction	Probably negligible	0	0	Probably negligible
Earnings of spouses, other income	Assumed negligible	0	0	Assumed negligible
Other government benefits	Assumed negligible	0	Assumed negligible	0

Notes: See Section 6.1 and Appendix G for discussion of each benefit/cost component. All benefits and costs are dollars per beneficiary over ten years and are inflation-adjusted to 2016 dollars and discounted to 2016 present value.

Unweighted sample sizes: T21 = 4,854, T22 = 3,041, C2 = 4,849.

A decline in voluntary work might have occurred had the BOND benefit offset substantially increased hours of paid work. Stage 2 impact findings for any paid work and for earnings show such small effects (at most $431 per year for earnings [Exhibit F-7]) that an important amount of displacement of volunteer work seems unlikely.[109]

Findings of net benefits to SSDI beneficiaries in both Stage 2 policy comparisons with current law might be somewhat magnified by taking account of impacts on health and life-satisfaction.[110] Although the per-year net benefits to Stage 2 treatment subjects of $764 (offset plus WIC) and $836 (offset plus EWIC) are considerably higher than the per-year net benefit of $158 to Stage 1 treatment subjects, these amounts are probably not high enough to have a substantial effect on health and life satisfaction.

Conceivably the offset could have affected beneficiary mortality, though the data show little indication of any effect on death rates. Mortality was 0.5 of a percentage point lower among T21 subjects receiving the offset plus WIC than among C2 subjects under current law (see Appendix F, Exhibit F-17). Prevalence in the T22 group receiving the offset plus EWIC was 1.0 percentage point lower than for the C2 group, and 0.5 of a percentage point lower than for the T21 group. None of these impact estimates differs statistically from zero. To the extent that the offset and/or EWIC counseling reduced deaths, that effect would manifest in the benefit-cost analysis as larger net benefits to SSDI beneficiaries (through more months of earnings and SSDI benefits and well-being gains from added months of life) and higher costs to the Disability Insurance Trust Fund. Society would gain through beneficiaries' expanded earnings and life-spans. However, we do not attempt to attach a monetary value to any of these potential benefits of additional life years. Were this done, the finding that could shift from a net cost to a net benefit is the value to society of EWIC counseling compared to WIC counseling: the current -$8,152 figure there (Exhibit 6-9) could even turn positive, if the mortality differential outweighed all other costs to society.

Another possible area of costs and benefits concerns the spouses of beneficiaries. It is possible that beneficiaries' marriage partners worked less or more because the earnings and SSDI benefits of their spouses changed. Compared to current law, increases in the incomes of T21 and T22 beneficiaries from

---

[109]   The offset's effect on voluntary work can be partially assessed by looking at the 36-month follow-up survey data. The estimated impact on hours of volunteer time per week was -0.1 and statistically insignificant for the offset plus WIC compared to current law, 0.0 for the offset plus EWIC compared to current law, and 0.1 for EWIC compared to WIC. Thus, if there was any impact on voluntary work for Stage 2, it was probably negligible.

[110]   There could have been substantial effects on the health and life-satisfaction of some of those who actually used the benefit offset, but because users comprise around 15 percent of the sample, the overall effects are likely to be small. The Stage 2 36-month follow-up survey asked about a number of self-reported health measures. Estimated impacts on these measures vary in sign for the Stage 2 policy comparisons and are generally of negligible size and statistically insignificant. The most illuminating of the measures asked respondents to say whether their health is excellent, very good, good, fair, or poor. There was no discernable movement along this scale, except that relative to the current law control group, slightly fewer treatment group members indicated that their health was excellent and slightly more that their health was poor, findings that are statistically insignificant.

spousal income were modest. [111] Still, of only 38 percent of Stage 2 volunteers are married or living with a partner (Gubits et al. 2013), limiting the potential for spousal earnings to change conclusions across all beneficiaries in the T21 or T22 groups. As noted earlier, we would not expect an average spousal earnings response that is larger than the small beneficiary income change due to the benefit offset, since most beneficiaries are not married. We impute the 36-month survey value of marital status to all years of the demonstration to characterize net costs and benefits in Appendix G (Exhibits G-10, G-11, and G-12). Those results are not qualitatively different from those presented here: the net benefit to society is positive for T21 compared to current law and negative for T22 compared to current law.

Due to the benefits offset policy, government transfer payments could potentially have changed in a variety of programs in addition to SSDI and SSI. Had the offset appreciably increased earnings and employment, for example, unemployment insurance payments might have increased, while payments under TANF and SNAP (Food Stamps) might have decreased. However, as previously noted, the Stage 2 BOND interventions had little effect on employment and earnings. Benefits from transfer programs such as TANF and SNAP may have declined due to increases in earnings and SSDI benefits but by well under a dollar for every dollar by which other income increased.[112] Moreover, many of those for which the offset increased earnings and/or SSDI benefits probably were not on TANF and SNAP given their access to these other income sources. [113] Further, the threat of reductions in TANF or SNAP benefits presumably would discourage treatment subjects enrolled in those programs from taking advantage of the benefit offset by increasing their earnings—making offset-induced reductions in TANF and SNAP benefits even less likely. In total, we judge the potential impact of the BOND interventions on other government benefits to be small and of uncertain direction.

Overall, as with Stage 1, we expect the conclusions of the net benefits from monetized components to be little changed when the benefits and costs of non-monetized components are considered.

6.5.   Robustness of the Base-Case Findings for Stage 2

As previously discussed, a number of limitations apply to the base-case benefit-cost results that appear in Exhibits 6-7, 6-8, and 6-9. To explore the importance of these limitations for the substantive conclusions above, we performed several sensitivity analyses. This section presents the results of those analyses.

---

[111]   The Stage 2 36-month survey provides a positive estimates of the offset's impact (with WIC or with EWIC compared to current law) on spouse's earnings, albeit small (around $400 [offset plus WIC] or $200 a year [offset plus EWIC]) and statistically insignificant ones. See Appendix G, Exhibits G-10 and G-11.

[112]   Based on the 36-month survey, the estimated impacts on payments under various government transfer programs for the three Stage 2 policy comparisons are small ($100 to $300 a year), of varying signs, and in no case statistically significant. This is not surprising because, except for SNAP, few SSDI beneficiaries participate in these programs. While almost a third of the Stage 2 subjects received SNAP benefits at the time of the survey, the estimated impacts on SNAP benefits were small ($25 to $35 a year) and statistically insignificant.

[113]   One-third (32 percent) of Stage 2 treatment group members indicated that they received SNAP benefits at the time of the 36-month survey. Only 1 percent received TANF.

6.5.1. Sensitivity to Projection Assumptions

One limitation for the 10-year accounting period is the assumptions that were made to project the benefits and costs observed during the demonstration five years into the future. Thus, we replicated Exhibits 6-7 through 6-9, but assumed zero benefits and costs in years 6 through 10. Stage 2 enrolled participants beginning in 2011. We use annual impacts on administrative earnings and benefit (SSDI and SSI) outcomes for 2012 to 2015. For all income sources, we impute one-fourth of the 2012 to 2015 aggregate impacts as the 2011 values.

These five-year findings follow a very similar pattern to those appearing in Exhibits 6-7 through 6-9. In each case, there are gains for beneficiaries and losses for the Disability Insurance Trust Fund, with society as a whole estimated to gain over a thousand dollars per beneficiary under offset plus WIC and lose four times that with offset plus EWIC. However, as would be anticipated, the monetized benefit and cost estimates are smaller when estimated over five years instead of over ten years, adding up to roughly half the net benefits and cost calculated for the 10-year accounting period.

Exhibit 6-10.   Stage 2 Net Benefits, by Perspective, for Alternative Projection Assumptions

		Beneficiaries	**DI Trust Fund**	**Other Government**	**All of Society**
T21 versus C2	5-year findings	$3,815	-$2,261	$439	$1,648
	10-year findings	$7,642	-$4,798	$882	$2,982
T22 versus C2	5-year findings	$4,214	-$5,819	$0	-$2,710
	10-year findings	$8,363	-$11,345	-$27	-$5,170
T22 versus T21	5-year findings	$399	-$3,563	-$440	-$4,365
	10-year findings	$721	-$6,547	-$909	-$8,152

Notes: All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W), and discounted to 2016 present value using a real discount rate of 2.7 percent.

Unweighted sample sizes: T21 = 4,854, T22 = 3,041, C2 = 4,849.

6.5.2. Distributional Weighting

Exhibit 6-11 shows results when we give a greater weight to beneficiary net incomes (discussed in Section 6.3.2). Using a distributional weight of 1.2 for beneficiaries, as shown in Exhibit 6-11, the net social value of the offset policy increases for both offset plus WIC and offset plus EWIC groups, but the difference between the two treatment groups diminishes slightly. However, there is no qualitative change in the comparison of WIC and EWIC when we give greater weight to beneficiary net incomes.

Exhibit 6-11.   Stage 2 Net Benefits, by Perspective, with Different Distributional Weights

	Beneficiaries	DI Trust Fund	Other Government	All of Society
Offset Plus WIC Compared To Current Law				
Base case, T21 vs C2	$7,642	-$4,798	$882	$2,982
Greater weight applied to beneficiary income	$9,170	-$4,798	$882	$4,510
Offset Plus EWIC Compared To Current Law				
Base case, T22 vs C2	$8,363	-$11,345	-$27	-$5,170
Greater weight applied to beneficiary income	$10,035	-$11,345	-$27	-$3,498
Offset Plus EWIC Compared to Offset Plus WIC				
Base case, T22 vs T21	$721	-$6,547	-$909	-$8,152
Greater weight applied to beneficiary income	$865	-$6,547	-$909	-$8,008

Notes: All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W), and discounted to 2016 present value using a real discount rate of 2.7 percent.

Unweighted sample sizes: T21 = 4,854, T22 = 3,041, C2 = 4,849.

6.5.3.   Halving or Doubling Uncertain Values

Relative to Stage 1, halving or doubling estimates other than SSDI impacts would have a larger impact in Stage 2. This is because estimated earnings impacts are larger, but changing assumptions on fringe rates or the value of non-market time by a factor of two would not change the sign of net benefits.

6.5.4.   Sensitivity to Deadweight Loss Assumption

As noted earlier, for the U.S. estimates of the METB range from 6 cents per dollar to 43 cents per dollar, with a median value of 19 cents per dollar. The estimates of deadweight loss in the base case in Exhibits 6-8, 6-9, and 6-10 use that median value of 19 cents per dollar. Were a value of 6 cents per dollar used instead, deadweight loss would shrink by 68 percent. Estimated net benefit to society for T21 (offset plus WIC) versus control (C2) would rise 17 percent (to $3,491). Estimated net cost to society for T22 (offset plus EWIC) versus control would fall 29 percent (to $3,692), but neither substantive conclusion would change. The signs of net benefit in each case would not change, because deadweight loss is small compared to other components of social costs. The measured disadvantage of the offset plus EWIC compared to the offset plus WIC would shrink 12 percent (to $7,183), assuming the METB were 6 cents per dollar, but the qualitative pattern of results would not change.

Similarly, if the METB were 43 cents per dollar, costs resulting from deadweight loss would grow by 126 percent. As a result, the net benefit to society for T21 versus control would fall by 32 percent (to $2,042). If the METB were 43 cents per dollar, the net cost to society for T22 versus control would rise by 53 percent respectively (to $7,900). In that case, the disadvantage of the offset plus EWIC compared to the offset plus WIC would rise 22 percent (to $9,942). Thus, the estimate of the cost to society is somewhat sensitive to the assumed value of the METB used to compute deadweight loss, but the conclusion on the sign of net benefits is unchanged in each case.

6.5.5. Partial Sensitivity Analysis

As in Stage 1, we assessed the sensitivity of our calculations to implementation challenges that would not apply in the steady state (Exhibit 6-12) and to different real discount rates (Exhibit 6-13). As we found in Stage 1, the results for net social impact are not qualitatively different as we vary these assumptions. However, if implementation had been complete prior to years 1 and 2 of the Stage 2 evaluation, so that a steady-state situation existed when the evaluation began, the benefits to beneficiaries would have been modestly larger and the costs to the Trust Fund would also have been modestly larger, with essentially offsetting impacts of roughly equal magnitudes.

Exhibit 6-12.   Stage 2 Net Benefits, by Perspective, Modeling Steady State

	Beneficiaries	DI Trust Fund	Other Government	All of Society
Offset Plus WIC Compared to Current Law				
Base case, T21 vs C2	$7,642	-$4,798	$882	$2,982
No implementation variation, T21 vs C2	$8,197	-$5,445	$949	$2,846
Offset Plus EWIC Compared to Current Law				
Base case, T22 vs C2	$8,363	-$11,345	-$27	-$5,170
No implementation variation, T22 vs C2	$8,886	-$11,840	-$57	-$5,271
Offset Plus EWIC Compared to Offset Plus WIC				
Base case, T22 vs T21	$721	-$6,547	-$909	-$8,152
No implementation variation, T22 vs T21	$689	-$6,394	-$1,005	-$8,117

Notes: The no implementation variation scenarios impute average impacts over 2014 and 2015 for the first three years of the demonstration. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W), and discounted to 2016 present value using a real discount rate of 2.7 percent.

Unweighted sample sizes: T21 = 4,854, T22 = 3,041, C2 = 4,849.

Exhibit 6-13 shows that varying the real discount rate has little qualitative impact on the estimated cost or benefit, because increased benefits in the form of higher earnings and benefits paid are realized in approximately the same time periods as increased costs in the form of benefits paid and administrative costs incurred. That is, there is not a large up-front gain followed by losses later on, or vice versa, so the discount rate has little impact on conclusions.

Exhibit 6-13.   Stage 2 Net Benefits, by Perspective, with Different Real Discount Rates

	Real Discount Rate	Beneficiaries	Disability Insurance Trust Fund	Other Government	All of Society
T21 vs C2	2.3%	$7,632	-$4,796	$879	$2,971
	2.7%	$7,642	-$4.798	$882	$2,982
	3.5%	$7,663	-$4,804	$890	$3,005
	5.5%	$7,731	-$4,829	$910	$3,069
T22 vs C2	2.3%	$8,350	-$11,323	-$29	-$5,158
	2.7%	$8,363	-$11,345	-$27	-$5,170
	3.5%	$8,390	-$11,394	-$23	-$5,196
	5.5%	$8,474	-$11,537	-$14	-$5,272
T22 vs T21	2.3%	$718	-$6,527	-$907	-$8,129
	2.7%	$721	-$6,547	-$909	-$8,152
	3.5%	$727	-$6,589	-$913	-$8,201
	5.5%	$742	-$6,709	-$924	-$8,341

Notes: All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W), and discounted to 2016 present value using various real discount rates (as indicated in the exhibit).

Unweighted sample sizes: T21 = 4,854, T22 = 3,041, C2 = 4,849.

6.5.6.   Monte Carlo Simulations

We conducted the Monte Carlo analysis for Stage 2 analogously to the analysis for Stage 1. Key findings from this analysis appear in Exhibit 6-14. As shown, the original estimates of total net benefits, and those derived by averaging the net gain values over the 2,000 Monte Carlo trials differ by much less than a standard deviation. This is unsurprising. Each Monte Carlo trial is based on random deviations from the original individual impact estimates.

Both the offset plus WIC (T21) and the offset plus EWIC (T22) compare favorably to current law (C2) when looking at beneficiaries, but not when considering the DI Trust Fund. These conclusions are robust to sampling error, in the sense that confidence intervals do not include zero, for either net benefits to beneficiaries or net costs to the DI Trust Fund. However, we cannot reject the hypothesis that there is no difference in net benefits to beneficiaries between the offset plus WIC (T21) and the offset plus EWIC (T22), as the confidence interval (-2,652 to 4,094) for the difference includes zero.

The offset plus WIC has a substantially smaller negative impact on the DI Trust Fund compared to current law than does the offset plus EWIC. The confidence interval for the difference in the two treatment arms indicates the offset plus EWIC would reduce the DI Trust Fund by roughly $3,000 to $10,000 per beneficiary over a 10-year horizon, compared to the offset plus WIC. This arises because of the large administrative costs associated with providing counseling to the offset plus EWIC group. The differences in counseling costs are due both to more expensive unit costs for EWIC compared to WIC and to higher take-up of counseling in the T22 group compared to the T21 group (96 percent compared to 38 percent).

The offset plus WIC compared to current law shows a positive net social benefit in the mean estimate, but the confidence interval includes zero. It follows that we cannot conclude the offset plus WIC had a positive net social benefit compared to current law. Comparing the offset plus EWIC to current law tends not to show a positive net social benefit, though the confidence interval for the benefit to all of society again includes zero. However, even though net benefits for both treatment arms have confidence intervals that include zero, comparing the two treatments shows that the offset plus EWIC would have worse consequences for net social benefit compared to the offset plus WIC, due to the larger costs to the DI Trust Fund.

It is a well-known, yet still surprising property of comparisons across treatment conditions that even when confidence intervals overlap, the confidence interval for the difference may not include zero. This arises because the variability of the difference is smaller. That explains why the findings can reject the null hypothesis that the offset plus WIC produces the same net social benefit as the offset plus EWIC even though the confidence interval for each includes zero.

Exhibit 6-14.   Sensitivity of Stage 2 Estimates to Sampling Variability

	Beneficiaries	Disability Insurance Trust Fund	Other Government	All of Society
Offset Plus WIC Compared to Current Law				
Base-case net benefits for T21 versus C2	$7,642	-$4,798	$882	$2,982
Mean net benefits from 2,000 trials for T21 versus C2	$7,498	-$4,462	$848	$3,197
Standard deviation for T21 versus C2	$1,320	$1,382	$558	$1,872
Confidence interval for T21 versus C2 (centered on base-case)	$5,055 to $10,228	-$7,506 to -$2,090	-$211 to $1,975	-$686 to $6,650
Offset Plus EWIC Compared to Current Law				
Base-case net benefits for T22 versus C2	$8,363	-$11,345	-$27	-$5,170
Mean net benefits from 2,000 trials for T22 versus C2	$8,257	-$11,516	-$47	-$5,503
Standard deviation for T22 versus C2	$1,663	$1,199	$946	$3,446
Confidence interval for T22 versus C2 (centered on base-case)	$5,102 to $11,623	-$13,696 to -$8,995	-$1,880 to $1,827	-$11,925 to $1,585
EWIC Instead of WIC, Given Offset				
Base-case net benefits for T22 versus T21	$721	-$6,547	-$909	-$8,152
Mean net benefits from 2,000 trials for T22 versus T21	$771	-$7,064	-$904	-$8,712
Standard deviation for T22 versus T21	$1,721	$1,735	$717	$2,285
Confidence interval for T22 versus T21 (centered on base-case)	-$2,652 to $4,094	-$9,947 to -$3,148	-$2,315 to $497	-$12,632 to -$3,673

Notes: The 95-percent confidence intervals are centered on the base-case net benefits. All dollar amounts are inflation-adjusted to 2016 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W), and discounted to 2016 present value using a real discount rate of 2.7 percent.

6.6.   Conclusions

The results of the benefit-cost analysis suggest that the main effect of the BOND benefit offset is to transfer more income from the Disability Insurance Trust Fund to SSDI beneficiaries than under current law. Beneficiary gains necessarily accrue to the relative few who are willing and able to work and thereby use the offset. So in this sense, the offset intervention resembles a work-promoting transfer program such as the Earned Income Tax Credit. Even so, dollar gains per person are small among the general beneficiary population (represented in Stage 1) because of the low offset utilization rate among them. More substantial gains occur on average for the small share of beneficiaries who volunteered for Stage 2 of the demonstration. The estimates show net social benefits for the offset plus WIC group (T21) as a result, but not for the offset plus EWIC group (T22) because of the relatively high costs of their counseling services.

One important difference between the Stage 1 and 2 findings concerns the net benefits of the offset policy to society as a whole: they are negative for Stage 1 but positive for one treatment group in Stage 2 (T21, offset plus WIC) and negative for the other (T22, offset plus EWIC).

The Stage 1 finding of negative net benefits to society is entirely due to an increase in the deadweight loss caused by the taxes presumed to be needed to fund increased SSDI benefits. The estimate of deadweight loss is based on the median estimate of marginal excess tax burden drawn from the economics literature from among a fairly wide range of estimates. Moreover, the negative Stage 1 social net benefit reverses to become a positive net benefit to society when allowance is made for the possibility that the marginal utility of income is larger for SSDI beneficiaries than the general population because of beneficiaries' relatively low incomes.

In contrast, net social benefits in Stage 2 are unaffected by the deadweight loss calculation, because deadweight loss is small compared to net social benefits. Net social benefits are positive and large for the offset combined with WIC, but negative and large in magnitude for the offset combined with EWIC. The difference is due to increased costs of counseling for T22 (the offset combined with EWIC) compared to T21 (the offset combined with WIC). For neither of these estimates can we rule out sampling variation as the cause of a nonzero estimate. Nevertheless, the difference between the two treatment arms is nonzero, in the sense that we can rule out sampling variation as the cause of that difference, so we conclude that T21 (the offset combined with WIC) has greater net social benefit than T22 (the offset combined with EWIC).

The findings that are briefly summarized in the previous three paragraphs were subjected to a number of different sensitivity tests, including the referenced Monte Carlo analyses. With the exception of the sensitivity to the assumed marginal excess tax burden already noted for Stage 1, the results appear to be remarkably robust.

7.   Interpretation and Implications

This chapter draws on all components of the BOND evaluation to interpret the findings and consider their implications. The chapter begins with a brief summary of the policy context for BOND, the findings from the pilot demonstration that preceded BOND, and a brief statement of BOND's main contribution to the evidence on how the SSDI earnings rules affect beneficiary behavior (Section 7.1). Section 7.2 provides additional evidence on the treatment subjects' behavioral responses to the BOND offset and considers why they were not larger. Then, Section 7.3 considers why the enhancements to work incentive counseling did not increase behavioral responses.

The chapter then considers the implications of BOND for future policy research. Section 7.4 takes advantage of BOND's two-stage design to learn more about the behavioral responses of the Stage 1 treatment subjects who would not have volunteered for Stage 2 had they been given the opportunity, and how they compare to the findings from Stage 2 volunteers who received the same treatment. The findings have important implications for making inferences about national policy from demonstrations that rely solely on informed volunteers. The section also considers the importance of timely benefit adjustments for understanding how incentives affect SSDI beneficiary behavior and the implications of BOND for other partial benefit designs of interest to policymakers.

7.1.   BOND's Main Findings in Context

SSDI is the nation's primary earnings replacement program for workers who become unable to work substantially due to long-term or terminal physical or mental conditions. Since the early 1990s, increasing numbers of SSDI claimants and low exit rates from the program have contributed to substantial and rapid growth in SSDI program costs. Rising program costs have led the Social Security Board of Trustees to project that the DI Trust Fund[114] will be exhausted by 2032 (SSA 2018), leading policymakers to consider changes to SSDI program rules that might reduce total SSDI benefits. Complementarily, low employment rates for people with disabilities have prompted policymakers to consider changes to SSDI program rules that might increase employment. More work among SSDI beneficiaries would lead to reduced benefits and higher payroll tax payments, which would be an aid in addressing the problems of sustaining the DI Trust Fund.

Observers widely presume that the "cash cliff" in current SSDI rules—the total loss of SSDI benefits after a sustained period of substantial earnings—discourages work. The Ticket to Work and Work Incentives Improvement Act of 1999[115] directed SSA to test a $1 for $2 benefit offset (a more gradual reduction in benefits in place of the cliff) to produce nationally representative estimates of the costs and benefits of an ongoing national benefit offset policy. A benefit offset policy will reduce the SSDI benefits of those induced by the offset to earn above the cash cliff level under current law. Nevertheless, for several

---

[114]   Social Security taxes and other income are deposited in the Disability Insurance (DI) Trust Fund and SSDI benefits are paid from it. Benefits are paid from the DI Trust Fund. The Annual Report of the Social Security Board of Trustees analyzes the actuarial status of the Old Age and Survivor's Insurance (OASI) and Disability Insurance (DI) Trust Funds (SSA 2018).

[115]   Ticket to Work and Work Incentives Improvement Act of 1999, Public Law 106-170, Section 302, 106th Congress, codified at U.S. Code 42 (1999) §434.

reasons, a benefit offset policy could raise program costs. First, a benefit offset policy would pay partial benefits to beneficiaries in those months when they would receive no benefits under current law—months in which their earnings exceed the cash cliff. That is, under the offset policy, SSA would pay partial benefits to those who would have had zero SSDI benefits under current law (in effect, a windfall to those who would already engage in substantial gainful activity under current law). Second, economic theory predicts that those beneficiaries receiving windfall income will respond by reducing their earnings somewhat, thereby further increasing average SSDI benefits. And third, a benefit offset policy would also increase program costs if it induced individuals to enter the SSDI program who would not do so under current law rules.

7.1.1.   What Does Previous Evidence Suggest About the Effects of a $1 for $2 Benefit Offset?

From 2004 to 2010, SSA tested a $1 for $2 benefit offset in the Four-State Benefit Offset Pilot Demonstration (BOPD) with the objective of testing procedures to administer this policy (Weathers and Hemmeter 2011). Implementation teams in four states recruited volunteers from a beneficiary group selected because its members were thought to be interested in increasing earnings. The specific groups varied by state. Each state's team randomly assigned consenting volunteers either to a treatment group that received the benefit offset or to a control group that continued to be subject to current law SSDI rules. Across all states, the pilot found that replacing the cash cliff with a $1 for $2 benefit offset increased the proportion of SSDI beneficiaries with earnings over the SGA threshold by 4 percentage points, or 25 percent of the control group mean. The pilot did not detect statistically significant impacts on average earnings, presumably because of two countervailing factors. Some beneficiaries earned more than SGA who would not otherwise have done so, which pushed average earnings up. However, other beneficiaries who would have earned more than the SGA threshold reduced their earnings, which pushed average earnings down. The BOPD also found that the benefit offset increased annual SSDI benefits by $544 per beneficiary in the second year after random assignment.

Though not nationally representative, the results of the pilot produced empirical evidence suggesting a benefit offset could lead to a significant increase in the proportion of beneficiaries with earnings over the SGA level. The pilot findings also suggested that a benefit offset might nonetheless increase benefit payments, even without induced entry, and also reduce the earnings of beneficiaries who would earn more than SGA under current program rules (Weathers and Hemmeter 2011).

The pilot also concluded that early problems administering the offset, stemming from the manual process SSA used to calculate benefits, may have affected the employment behavior of some beneficiaries. For example, some beneficiaries received notices with incorrect information about their SSDI benefits. In some cases, errors applying the offset rules led to under- and overpayments to beneficiaries (Weathers and Hemmeter 2011; Tremblay et al. 2011; Chambless et al. 2011). The pilot produced recommendations to automate and improve the administrative procedures used to adjust benefits according to benefit offset rules. The pilot's results underscored the importance of work incentives counseling to inform beneficiaries about how the offset rules work.

7.1.2.   Why Did SSA Conduct BOND?

SSA implemented the Benefit Offset National Demonstration (BOND) to develop nationally representative estimates of the costs and benefits of a $1 for $2 benefit offset.[116] BOND includes more than 90,000 treatment subjects and more than 900,000 control subjects who reside in a randomly-selected, nationally-representative set of 10 sites. BOND is the largest demonstration that SSA has conducted in its efforts to test policies that have the potential to help SSDI beneficiaries increase work and earnings and reduce SSDI benefits due to beneficiaries.

BOND's design incorporated lessons from the 4-state pilot regarding demonstration operations. Specifically, SSA developed an automated stand-alone data system to calculate benefits according to the benefit offset rules with the objective of increasing timeliness and accuracy of benefit adjustments. SSA also attempted to expedite work CDR completion for beneficiaries by having demonstration staff assist in documenting beneficiaries' past work. In addition, BOND tested enhancements to work incentives counseling that were developed in response to lessons from the pilot regarding beneficiary confidence in and knowledge of offset rules.

7.1.3.   How Does BOND Contribute to the Evidence About the Impacts of a National Benefit Offset?

The evidence from BOND shows that a $1 for $2 offset starting at the annual equivalent of SGA, coupled with minimal modifications to work-incentive counseling, will not produce large enough effects on beneficiary earnings to reduce government expenditures for the support of workers with disabilities. As we will consider further in the next section, there are reasons to think that the behavioral impacts of the benefit offset under a national policy might be larger than the Stage 1 estimates would imply. There is considerable uncertainty about whether this would be the case, however. The combined evidence from Stages 1 and 2 leaves no reason to think that earnings impacts could be large enough to reduce net government transfer payments to disabled workers. These results hold even before considering potential expansions in government expenditures due to induced entry. The BOND results also show that enhancements to counseling like those tested in Stage 2 do not increase beneficiaries' responses to the offset's work incentive.

The benefit-cost analysis found a net cost to society (i.e., costs exceed benefits) of the BOND offset for the full SSDI caseload examined in Stage 1. The very small estimated increases in earnings were not sufficient to offset the deadweight loss from increases in taxes to fund larger transfer payments. Distributional effects were much larger, with SSDI beneficiaries gaining income by receiving larger SSDI benefits and countervailing losses occurring for the Disability Insurance Trust Fund. The benefit-cost analysis for Stage 2 shows that the offset policy combined with standard work incentives counseling has a net benefit to society in the small subpopulation of beneficiaries who volunteer for the demonstration. In

---

[116]   After considerable consultation with experts, SSA determined that it would not be feasible for BOND to measure induced entry using an experimental research design (Tuma 2001). Nonetheless, induced entry remains an important issue and SSA plans to assess the likely extent of induced entry through separate non-experimental research.

contrast, the offset policy combined with enhanced work counseling has a net social loss, a result largely due to higher counseling costs.[117]

7.2. The Earnings Impacts from a National Offset Policy Would Not be Large Enough to Reduce Net Government Expenditures for Disabled Workers

This section begins by explaining how the evidence from BOND leads to the conclusion that the earnings impacts from a national offset policy would not be large enough to reduce net government expenditures for disabled workers. The section considers Stage 1 evidence first, then describes how the Stage 2 evidence increases our confidence about that conclusion. Finally, the section considers potential reasons why the impacts on earnings were not larger.

7.2.1. Earnings and Benefits Effects for Stage 1 Beneficiaries

Evidence from Stage 1's nationally representative sample of SSDI beneficiaries shows that a national benefit offset policy would not increase the share of beneficiaries engaged in substantial work by enough to reduce net government expenditures on SSDI benefits. The point estimate for the impact on earnings over five years is essentially zero (Exhibit 5-2). The 95 percent confidence interval for that effect ranges from -3.0 percent to +3.2 percent of earnings. The point estimate for the impact on SSDI benefits is $143 per year per beneficiary, with a 95 percent confidence interval that ranges from 0.9 percent to 1.6 percent of current law benefits (Exhibit 5-2). We discuss the findings for these two variables further, below.

Discussion of Earnings Impacts
As we discuss below, there are reasons to think that impacts on the proportion of beneficiaries earning above BYA might be somewhat larger in a (non-demonstration) national program than what was observed in the nationally-representative Stage 1 sample. However, further examination of the Stage 1 earnings impacts suggests that the impact of a national policy on the percentage with earnings above BYA would have to be 31 times larger than the corresponding Stage 1 impact estimate to produce a breakeven outcome of zero impact on average SSDI benefits ("benefit neutrality"). That the Stage 1 impact could be 31 times larger strains credulity. In particular, 6.2 percent of beneficiaries would need to be induced to earn more than BYA, compared to the 0.2 percent point estimate for Stage 1. This 6.2 percent figure is derived from the Stage 1 estimates for 2014—late enough in the demonstration period so that various factors that might have delayed a behavioral response had substantially subsided and early enough so that future adjustments to SSDI benefits data would not likely lead to a substantive revision to the finding.[118]

---

[117]   The net benefit to society from the offset plus WIC compared to current law results from a 9 percent earnings increase that more than pays for the deadweight loss of higher taxes to fund increased transfer payments. The net loss to society from the offset plus EWIC compared to current law is largely driven by high costs of counseling. The use of volunteers in Stage 2 means that this finding is not directly relevant to the benefits and costs of a benefit offset applied to the entire SSDI caseload. (Again, this analysis does not consider the benefits and costs of induced entry, which would increase deadweight loss and hence costs to society.)

[118]   Appendix H describes the decomposition in detail. The identifying assumptions are: 1) mean monthly benefits under the offset to those in each treatment group who would have received no benefits under current law (hereafter the "windfall" group) are equal to mean monthly benefits for those induced to earn more than BYA (hereafter the "increased earnings" group); and 2) mean full monthly benefits for both groups are the same as for all control subjects. These assumptions are sufficient to decompose the impact on average benefits into the

The 31-fold increase should be treated as only a rough indicator of how much larger the impact on the percentage earning more than BYA would need to be for the national policy to be benefit neutral. Both sampling errors and departures from the modest assumptions used in the multiple's calculation imply that the estimate is only rough.

However, there are two clear reasons to expect the required multiple to be of this order of magnitude. The first reason is that the percentage of months in which Stage 1 treatment subjects received a benefit windfall in 2014 is more than 8 times as large as the percentage of treatment subjects induced to earn above BYA in that year: 1.67 percent versus 0.20 percent.[119] The second reason is that the average monthly reduction in benefits to those induced to earn more than BYA appears to be small relative to the average monthly SSDI benefits in windfall months.

To understand why the second reason is important, first consider the case where all offset users are receiving partial benefits that are exactly half of their full benefit. Under this scenario, the savings from one extra beneficiary induced to earn over BYA (a reduction from full to half benefits) is exactly equal to the extra expenditure for one beneficiary in the windfall group (an increase from zero to half benefits). Relative to this case, moving "down" the ramp (increasing earnings further from BYA) means that savings will be greater than windfall. Moving "up" the ramp (decreasing earnings closer to BYA) means that windfall will be greater than savings. Therefore, understanding where offset users are located on the ramp is important when considering how much larger effects would need to be to achieve benefit neutrality.

While we cannot directly observe the mean benefit reductions for those induced to earn more than BYA or mean benefits in windfall months, we do have information about the earnings distributions in the treatment and control groups in 2015. The earnings distributions for 2015 appear in Exhibit 7-1, along with the estimated impacts on the percentage in each earnings range; the impacts are also depicted in Exhibit 7-2.

Because the average annual benefit is very close to BYA, one would expect the bulk of offset users with partial benefits to have earnings between BYA and 3xBYA (rather than more than 3xBYA). Although some beneficiaries with earnings more than 3xBYA would be eligible for a partial benefit, most would not because the average full benefit is just slightly larger than BYA. (If full benefits were exactly equal to

---

positive impact because of partial benefits due to those in the windfall group and negative impacts for those in the increased earnings group. The size of the increased earnings group necessary to yield no impact on average benefits is conditional on no change in the size of the windfall group and no change in mean benefits for those in both groups. Although the identifying assumptions are not exactly correct, the findings are also not very sensitive to small departures from either assumption.

[119] Technically, the 1.67 percent is the percentage of person-months across the sample when Stage 1 treatment subjects received a benefit windfall (i.e., when these subjects were due a positive amount of SSDI benefits under the offset rules when under current law their benefits would have been zero). This percentage is calculated from 2014 results in Exhibit F-2: the impact on average number of months with SSDI benefits divided by the C1 mean is 0.20 months/10.35 months = 1.67 percent. The percentage of treatment subjects induced to earn over BYA (Exhibit F-1, 2014 impact on proportion with earnings above BYA is 0.20 percentage points) may also be considered a percentage of person-months, making these percentages directly comparable.

BYA, the offset would reduce them to zero at exactly 3xBYA.) The distributions shown in Exhibit 7-1 imply that 48 percent of control subjects and 52 percent of treatment subjects with earnings between BYA and 3xBYA have annual earnings between BYA and 1.5xBYA. Under the offset rules, their benefits would be reduced by at most 25 percent. Another 23 percent of control subjects and 24 percent of treatment subjects in the BYA to 3xBYA range have earnings between 1.5xBYA and 2xBYA, meaning that the offset would reduce their benefits by no more than 50 percent.[120] The impact estimates (Exhibit 7-2) also show that most treatment subjects induced to change their earnings by the offset (whether positive or negative) had earnings between BYA and 2xBYA. Therefore, these earnings distributions demonstrate that the bulk of those who would receive a partial benefit under the offset, whether increasing earnings above BYA or because they are recipients of a benefit windfall, have earnings that are located on the top half of the ramp, between BYA and 2xBYA. These earnings levels would result in only modest benefit reductions under the offset, implying only modest savings from those induced to earn above BYA, and relatively large windfall payments (that are more than half of full benefits).

The estimates for the impacts on the earnings distribution have another interesting implication. Contrary to expectations, the offset induced very few of those who would have kept their 2015 earnings just below the SGA amount under current law to avoid benefit loss (often referred to as "parkers") to increase their earnings to above BYA. Presumably most parkers would have 2015 earnings in the 0.5xBYA to BYA range under current law, so if the offset induced many to earn more than BYA we would expect to see a reduction in the percentage with earnings in that range. Out of every 10,000 control beneficiaries in 2015, only 350 had earnings in the 0.5xBYA to BYA range. And the non-statistically significant point estimate for that range, -0.05, implies that only 5 were induced by the offset to move to a different range— presumably to above BYA.[121] The point estimates imply, however, that 27 of 10,000 beneficiaries were induced to earn more than BYA.[122] It may be that more than 5 parkers moved from the 0.5xBYA to BYA range to a range above BYA, with others who would have had earnings below 0.5xBYA moving into the 0.5xBYA to BYA range to take their place, but there is no reason to expect the offset to encourage movement from below this band into this band. Instead, it is more plausible that the 26 in 10,000 induced to move from zero earnings to positive earnings constitute a large majority of the 27 induced to earn more than BYA.

An important caveat to this finding is that it is for 2015 only, and ignores the dynamics of earnings. If we counted as parkers all those who under current law would have had earnings between 0.5 BYA and BYA

---

[120] These benefit reductions under the offset rules may overstate the actual reductions because they do not take into account IRWE and other non-countable earnings or the much higher value of BYA for blind beneficiaries. Note, however, that the impact analysis found a modestly negative impact on average non-countable earnings.

[121] Statistics presented in Section 4.3 are consistent with the finding that few beneficiaries induced to increase earnings above BYA would have been parkers under current law. Specifically, if beneficiaries intentionally limit earnings to avoid complete benefit suspension, we would expect those beneficiaries to quickly increase earnings and maintain high earnings when the offset became available. However, beneficiaries who used the offset in 2011 or 2012 and continuously thereafter constitute only 25 percent of all offset users to date.

[122] The estimate of 27 in 10,000 induced to earn above BYA is based on summing the point estimates for the three below-BYA earnings categories. This characterization of movements across earnings ranges is based on Stage 1 point estimates. Movements across earnings categories under a permanent offset policy might follow a similar pattern, but with different relative magnitudes.

in at least one year of the 2011-2015 period, they would account for a larger share of those induced to earn above BYA than do those who would have had 2015 earnings between 0.5 BYA and BYA under current law. This broader definition of parkers is of relevance because the offset may affect behavior across multiple years. For instance, among control subjects with no earnings in 2015 who had earnings between 0.5 BYA and BYA in earlier years, there are likely some who would have earned above BYA in 2015 had they been assigned to T1. A finding from analysis of T1 impacts on earnings above BYA in at least one year by employment status in 2010 is consistent with our expectation that a multi-year definition of parking under current law, rather than a single year definition, would imply that parkers constitute a larger share of those induced to earn above BYA in any single year. Of those induced to earn above BYA in at least one year from 2011-2015, 60.9 percent were employed in 2010.[123]

Discussion of Benefit Impacts

The estimated impacts on SSDI benefits represent the bulk of the impacts of the BOND offset on DI Trust Fund benefit expenditures, but do not represent all of the impacts, for two reasons. The first additional impact on the Trust Fund expenditures is retroactive adjustments made after May 2017. We expect these to be quite small, and cannot predict whether they will ultimately have positive or negative effects on Trust Fund expenditures for the period.

The second additional impact on Trust Fund expenditures is due to the impact on unrecovered overpayments. This impact seems likely to be negative, and is potentially substantial, but will also likely be small in comparison to reported impact estimates for SSDI benefits. Unrecovered overpayments represent a Trust Fund benefit expenditure that is not captured in the SSDI benefit measure. We expect a negative impact on unrecovered overpayments for two reasons. First, the impact analysis of overpayments found a negative impact on mean overpayments, so the amount of overpayments to be recovered is lower. Second, SSA can more readily recover overpayments made to offset users than they can from those whose benefits are suspended or terminated for work under current law; the agency can reduce their subsequent partial payments.

It is not feasible to measure the impact on unrecovered overpayments because, in practice, it may take many years until the recovery of overpayments is completed and the extent of unrecovered overpayments are known. Some beneficiaries have repayment plans that extend for many years, in some cases past 2049 (SSA Office of the Inspector General [OIG] 2017), and it is unclear in advance whether beneficiaries will meet all of their repayment obligations. One recent analysis found that 47 percent of overpayments were unrecovered 10 years after the overpayment was established (SSA OIG 2015).

We can, however, put reasonable bounds on the size of the impacts on unrecovered overpayments. Consider Stage 1 first. We estimated a negative impact of $98 on mean overpayments over the whole

---

[123]   Appendix F reports that 14,688 T1 subjects were employed in 2010 and 62,413 were not employed (Exhibit F-20). The point estimates of impacts on the percentage with earnings above BYA in at least one year of 2011-2015 are 1.26 and 0.19 percentage points, respectively. These estimates imply that the number of T1 beneficiaries induced to earn more than BYA from the two T1 subgroups are 185 and 119, respectively. Although we find that those employed in 2010 account for 60.9 percent of those induced to earn above BYA in the demonstration, the corresponding ratio in a national program could be quite different because the standard errors for these estimates are substantial (equivalent to 78 and 44 subjects, respectively).

evaluation period (Exhibit 4-7) and a positive impact of $665 on benefits (Exhibit 5-2).[124] If we assume that 50 percent of overpayments for both groups will never be recovered—roughly in line with the OIG estimate for 10 years—the size of the impact on unrecovered overpayments will be -$49, so the size of the impact of the offset on Trust Fund benefit expenditures will be about 7 percent lower than the estimated impact on benefits. The effect would be more negative, and more substantial relative to the estimated impact on benefits, if the rate of unrecovered overpayments is lower for the T1 subjects. For instance, if SSA ultimately recovers 75 percent of the T1 group's overpayments, but only 50 percent of those for the control group, the impact on unrecovered overpayments would be -$112.[125]

The corresponding calculations for Stage 2 produce the following results: impacts on unrecovered overpayments of the combined treatment subjects (T21 and T22) is -$138.50 under an assumption of 50 percent recovery for all groups—about 8 percent of the estimated impact on T21 benefits ($1,791) and about 7 percent of the estimated impact on T22 benefits ($1,997) (Exhibit 5-5). If we instead assume 50 percent recovery for the control group but 75 percent recovery for each treatment group, the implied impact estimates for unrecovered overpayments for the combined Stage 2 treatment subjects is -$327.[126]

In summary, the estimated impact of the offset on benefits, as measured for this report, represents the bulk of the effects of the offset on DI Trust Fund benefit expenditures. This estimate likely overstates the impact on Trust Fund expenditures by a modest amount, however, because it appears that the offset would reduce unrecovered overpayments. Based on what we have learned about overpayments, the negative impact on unrecovered overpayments seems likely to be modest relative to the impact on SSDI benefits, but it is not feasible to predict that impact precisely because SSA pursues recovery of overpayments over many years.

7.2.2. Earnings and Benefits Effects for Stage 2 Beneficiaries

The impact estimates for Stage 2 treatment subjects also make it seem unlikely that a national offset policy would have sufficient impacts on beneficiary earnings to be benefit neutral. Compared to Stage 1 subjects, the Stage 2 treatment subjects' knowledge and confidence about how the offset would affect their benefits is likely to be closer to what knowledge and confidence would be under a national policy. (Indeed, Stage 2 treatment subjects' understanding is closer to control subjects' understanding of current law, see section 7.2.3.) Even so, the Stage 2 impact on the share with earnings above BYA—about 23 percent of the control group proportion—was not sufficient to reduce average SSDI benefits. The positive point estimates for SSDI benefits is equal to about 4 percent of the control group mean for both Stage 2 treatment groups, with the 95 percent confidence interval ranging from 2 to 6 percent.

---

[124] Slightly different samples were used for the overpayment analysis, but it seems likely that the impact on mean overpayments would be only slightly different—most likely smaller—had we been able to use the full sample.

[125] Mean overpayments for T1 subjects were $250 over the five years (Exhibit 4-7), of which 25 percent is $62.50. Mean overpayments for C1 subjects were $348, of which 50 percent is $174. The difference is $111.50.

[126] Mean overpayments for T21 and T22 subjects, combined, were $753 over the five years (Exhibit 4-8), of which 25 percent is $188.25. Mean overpayments for C2 subjects were $1,030, of which 50 percent is $515. The difference is $326.75.

For Stage 2, we find that the percentage of treatment subjects induced by the offset to earn more than BYA would have to be nearly 10 times larger than the actual measured impact to achieve benefit neutrality.[127] As with the 30-fold figure for Stage 1, this is an order-of-magnitude estimate. The two basic facts underlying the estimate are similar to those for Stage 1. First, in 2014, the increase in person-months in which SSDI benefits are due (i.e., the windfall months) is about twice as large as the percentage of person-months in which treatment subjects are induced to earn more than BYA (i.e., the "savings" months).[128] Second, as with Stage 1, the 2015 earnings distributions for treatment and control subjects (Exhibit 7-3) imply that the bulk of the Stage 2 treatment subjects most likely to have received partial benefits under the offset were due benefits that were only modestly lower than full current law benefits—including those induced by the offset to change their earnings (Exhibit 7-4).

Compared to the Stage 1 estimates, the Stage 2 estimates show more evidence that the offset induced some parkers to earn more than BYA, especially among those in the T21 group (Exhibit 7-4). For that group, the point estimates imply that somewhat more than half of T21 subjects induced to earn more than BYA would have earned between 0.5xBYA and BYA under current law, with most others having no earnings at all. For the T22 subjects, the point estimates imply that only 19 percent of those induced to earn more than BYA would have earned in the 0.5xBYA to BYA under current law, with almost all of the others having no earnings under current law. As noted above in the discussion of parkers among T1 subjects induced to earn more than BYA under the offset rules, a multi-year definition of parking under current law, rather than a single year definition, would imply that parkers constitute a larger share of those induced to earn above BYA in any single year.[129]

---

[127] This figure is based on the same methodology used to produce the corresponding figure for Stage 1. See Appendix H for the details. We derived separate multiples for the T21 group (7.6) and the T22 group (9.6).

[128] In 2014, for windfall benefits were due to T21 subjects in 5.2 percent of months, and 2.6 percent of T21 subjects were induced to earn more than BYA. The corresponding percentages for T22 subjects are 5.2 percent and 2.7 percent, respectively.

[129] The Stage 2 findings for impacts on earnings above BYA in at least one year by baseline employment status are of relevance to this point. Based on the findings for the combined Stage 2 treatment groups shown in Appendix Exhibit F-32, we know that 47.6 percent of those induced to earn above BYA in at least one year were employed at baseline. Of all Stage 2 treatment subjects, 1,909 were employed at baseline and 5,927 were not. The respective point estimates of the impact on the percentage with earnings above BYA in at least one year of 2012-2015 are 7.37 and 2.61 percentage points. These estimates imply that the number of beneficiaries induced to earn more than BYA from the two groups are 141 and 155, respectively. Although we find that those employed at baseline account for just under half of those induced to earn above BYA in the demonstration, the corresponding ratio in a national program could be quite different because the standard errors for these estimates are substantial (equivalent to 37 and 43 subjects, respectively).

Exhibit 7-1.   Estimated Impacts on 2015 Earnings Ranges: T1 Versus C1

2015 Earnings Range	Percent of T1 Group with Earnings in Range	Percent of C1 Group with Earnings in Range	Impact Estimate in Percentage Points	Standard Error
Earnings ($ in year)				
$0	86.75	87.01	-0.26*	0.13
$1 to 0.5x BYA	6.57	6.53	0.04	0.14
0.5x BYA to BYA	3.44	3.49	-0.05	0.08
BYA to 1.5x BYA	1.31	1.03	0.29***	0.05
1.5x BYA to 2x BYA	0.61	0.50	0.11***	0.03
2x BYA to 2.5x BYA	0.33	0.36	-0.03	0.02
2.5x BYA to 3x BYA	0.27	0.28	-0.01	0.02
Above 3x BYA	0.71	0.80	-0.09**	0.04

Source: SSA administrative records from the MEF and MBR.

Notes: Earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Chapter 2 for further detail). Weights ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics. The BOND Yearly Amount (BYA) is 12 times the monthly threshold for SGA. For non-blind beneficiaries, the 2015 BYA amount is $13,080. For blind beneficiaries, the 2015 BYA amount is $21,840.

Unweighted sample sizes: T1 = 77,101; C1 = 891,429.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test with 9 degrees of freedom (resulting from a research design involving 10 study sites) and with no multiple- comparisons adjustment.

Exhibit 7-2.   Graphical Depiction of Impacts on 2015 Earnings Ranges: T1 vs. C1
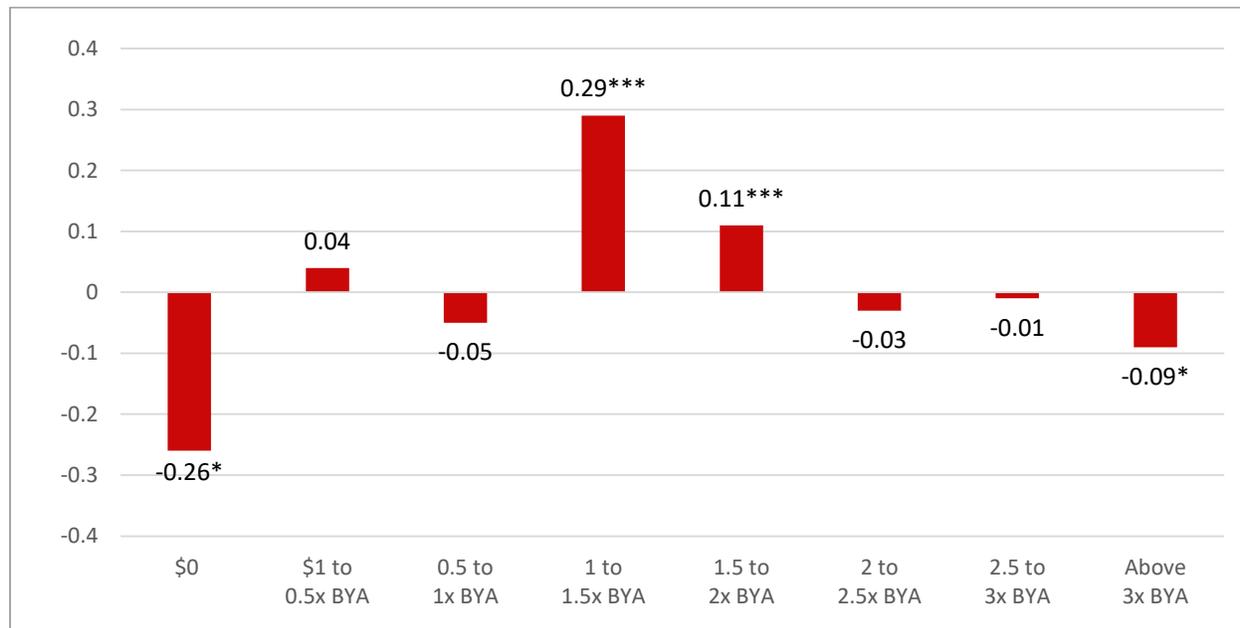
Exhibit 7-3.   **Estimated Impacts on 2015 Annual Earnings Ranges for Stage 2 Volunteers in Percentage Points: All Policy Comparisons**

Outcome	Average Outcome with Offset and WIC (T21)	Average Outcome with Offset and EWIC (T22)	Average Outcome under Current Law (C2)	Estimated Impact of Offset + WIC vs Current Law (T21 vs. C2)	T21 vs C2 SE	Estimated Impact of Offset + EWIC vs Current Law (T22 vs. C2)	T22 vs C2 SE	Estimated Impact of EWIC instead of WIC Given Offset (T22 vs. T21)	T22 vs T21 SE
$0	63.23	62.05	64.41	-1.18	(1.06)	-2.36*	(1.14)	-1.18	(1.15)
$1 TO 0.5x BYA	14.11	14.15	14.20	-0.09	(0.81)	-0.05	(1.30)	0.04	(1.00)
0.5x BYA to 1x BYA	10.19	10.97	11.54	-1.34	(1.07)	-0.56	(0.83)	0.78	(0.81)
1x BYA to 1.5x BYA	4.61	5.61	3.59	1.02*	(0.50)	2.02***	(0.56)	1.00	(0.60)
1.5x BYA to 2x BYA	2.78	2.47	1.66	1.12**	(0.37)	0.81*	(0.41)	-0.31	(0.47)
2x BYA to 2.5x BYA	1.71	1.84	1.30	0.41	(0.30)	0.54	(0.32)	0.14	(0.36)
2.5x BYA to 3x BYA	1.13	0.73	1.03	0.10	(0.26)	-0.29	(0.25)	-0.40	(0.29)
Above 3x BYA	2.23	2.16	2.27	-0.04	(0.38)	-0.11	(0.43)	-0.07	(0.37)

Source: SSA administrative records from the MEF and MBR and the Stage 2 Baseline Survey.

Notes: All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Chapter 2 for further detail). Weights ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics. All dollar values in 2016 dollars.

Unweighted sample sizes: T21 = 4,854 ,T22 = 3,041 , C2 = 4,849

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test with 9 degrees of freedom (resulting from a research design involving 10 study sites) and with no multiple- comparisons adjustment.

Exhibit 7-4.   Graphical Depiction of Impacts on 2015 Earnings Ranges for Stage 2 Volunteers: Comparisons With Current Law
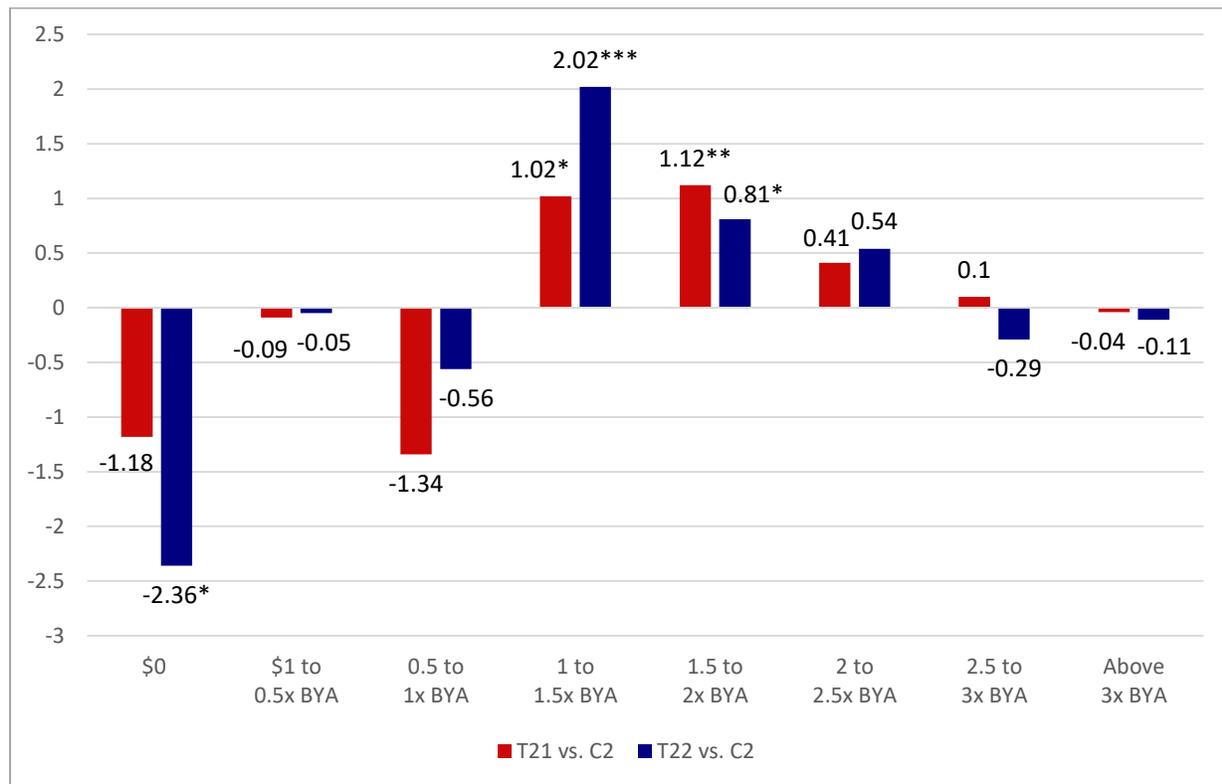
7.2.3. Why Impacts on Earnings Were Not Larger

There are at least four potential explanations for why offset use and the effect on percentage with earnings above BYA were not larger in magnitude.

1. One possible explanation is the limited *work capacity* of most beneficiaries. Because of the nature of the SSDI eligibility criteria, only a fairly small minority of beneficiaries were expected to use the offset.

2. A second possible explanation is that the increase in the *incentive* to earn more than BYA, although strong, was not strong enough to induce some beneficiaries with the capacity to earn more than BYA to actually do so. The offset imposes an implicit 50 percent tax on earnings. For some, this implicit tax, perhaps along with potential tax increases and reductions in other benefits, may make the increase in net income from increased earnings smaller than the opportunity cost of giving up other activities.

3. A third possible explanation stems from the *complexity* of both the intervention and the current law rules. The complexity of the rules may have muddled treatment subject understanding of the change in incentive available through the offset.

4. The final possible explanation interacts with the complexity of the rules: that the conditions established by the BOND *implementation* may have led to smaller impacts than would be the case under a permanent national benefit offset policy.

The balance of this section considers each of these four explanations.

Work Capacity

Because all SSDI beneficiaries were determined by SSA to have sufficiently severe impairments that they were unable to engage in SGA for at least 12 months, there is no expectation that a majority or even a substantial minority would be able to engage in SGA, whatever the earnings rules. Consistent with SSA's eligibility screening, when asked about their health during the 36-month survey, only 30 percent said their health was good, very good, or excellent.

Some beneficiaries who are currently unable to engage in SGA might be able to do so with the assistance of rehabilitation, training, and employment services. Although treatment subjects remained eligible for the same rehabilitation, training, and employment services available to all SSDI beneficiaries, BOND did not offer any additional services in conjunction with the benefit offset. However WIC and EWIC providers referred subjects to these services and subjects could seek these services on their own. The findings from the BOND participation and process analyses give some insight into the role that services of this type might have played in beneficiaries' attempts to build work capacity.

WIC and EWIC counselors reported that, for some beneficiaries, lack of access to employment services posed challenges to working and using the offset (see Chapter 4 and Appendix C). This finding raises the possibility that the lack of availability of services may have prevented some treatment subjects from increasing their capacity to work. As we show below, however, the evidence on service use and self-reported unmet needs for employment services does not support this interpretation.

The evaluation did not find an impact of the offset on use of rehabilitation, training, and employment services for Stage 1. If lack of availability of services had prevented the emergence of a differential in service use, we would expect to see greater unmet need for employment supports among treatment subjects. However, the 36-month survey responses provide no evidence that the offset affected self-reported unmet needs for employment supports (rates for T1 and C1 subjects were both approximately 37 percent).

Relative to current law and to the offset plus WIC, the offset plus EWIC increased use of rehabilitation, training, and employment services, presumably because EWIC offered additional referrals relative to WIC. Still, the increases were modest, about 6 percentage points for both comparisons, (for use of EN services, relative to 20 percent in T21 and 19 percent in C2). If this greater use of services for the EWIC group was successful in increasing capacity to work, we would have expected to observe fewer unmet needs for employment supports in the offset-plus-EWIC group than the other two groups. However, beneficiaries in all three Stage 2 groups reported unmet needs for services at the same level, about 49 percent. The equivalence in unmet needs between the groups is consistent with the interpretation that existing services that beneficiaries are accessing are not making a meaningful difference in work capacity, or at least self-perceived work capacity.

The Offset's Incentive

Replacing the cash cliff with a ramp for beneficiaries who engage in SGA substantially decreases the disincentive to earn more than BYA. Beneficiaries not already engaging in SGA would have to increase their work effort to take advantage of the benefit offset. Working more would require giving up time spent on other activities, and for some the implicit cost of giving up those activities—the opportunity cost—may be too high, despite the more favorable SSDI earnings rules.

The $1 for $2 offset for earnings above BYA is the equivalent of a 50 percent marginal tax rate on earnings between BYA and the level of earnings at which benefits fall to zero. Taxes on earnings and implicit taxes from reductions in other benefits (for example, SNAP and private disability benefits) further reduce the incentive to earn more than BYA under the offset. Hence, a beneficiary who values the activities that would be displaced at 50 percent or more of the potential increase in gross earnings from working more would presumably not choose to earn more. Alternative activities that might be more attractive than SGA-level work under both current law and the offset might include various forms of unpaid work, including volunteer work, caring for child or other relative, and any other work for which compensation is in-kind or informal.

The Complexity of the Earnings Rules

Current law rules are very complex. The offset policy replaced a fairly simple component of the current law rules with a more nuanced rule. The cash cliff is in effect, an on-off benefit switch, in which a beneficiary either receives her full benefit amount (if countable earnings are less than the SGA level after completing the trial work period and grace period), or zero benefit (if countable earnings exceed SGA after the trial work period and grace period). The ramp tested in BOND is more like a dimmer switch. In BOND, when countable earnings exceed BYA, monthly SSDI benefits could be any amount from full benefits to $0, depending on earnings. Under both sets of rules, a beneficiary might be confused about when SSA will take action to change SSDI benefits. However, knowing what will happen to benefits when SSA takes action is clearer under current law (suspend benefit payments) than under the offset (reduce benefit payments by some amount).

The BOND evaluation produced evidence indicating limited understanding, perhaps due to the complexity of the rules. Among Stage 1 treatment subjects, three years into the demonstration just 29 percent gave responses consistent with a correct understanding of how increased earnings affect SSDI benefits under the offset (Chapter 3). By comparison, 54 percent of Stage 1 control subjects gave responses to survey questions consistent with a correct understanding of how earnings affect benefits under current law. From this differential alone, it is unclear what to expect in a national offset program. It could be that under an ongoing national program, understanding of the offset rules would be about the same as the C1 subjects' understanding of current law. However, other interpretations are possible. Perhaps, the additional nuance of the offset rules would mean that understanding would never rise to the level seen in the C1 group.

The evidence from Stage 2 provides some additional insight. About half (48 percent) of T21 subjects gave survey responses indicating a correct understanding that, under the offset rules, increased earnings would cause benefits to drop but not to zero. This percentage is greater than the 29 percent found in the T1 group. It could be that the higher understanding of the T21 group compared to the T1 group is due to a better ability to grasp rules among volunteers than among the general beneficiary population. Or it could be that the more extensive information provided in Stage 2 to recruit volunteers led to higher understanding of Stage 2 treatment subjects. The level of correct understanding among the C2 group—

which at 54 percent is exactly the same as the C1 group—leads us to suspect that the additional information provided during volunteer recruitment is the main factor driving the difference in understanding between T21 and T1 subjects. The fact that the C2 group has no better understanding of current law than the C1 group does not suggest a greater ability to grasp rules on the part of volunteers.

Compared to the standard WIC, the more intensive EWIC increased the proportion of volunteers with correct understanding of the offset rules from 48 percent to 52 percent. This increase is statistically significant, but small. Relative to offset-plus-WIC subjects, offset-plus-EWIC subjects were more likely to have contact with a counselor (95 percent for EWIC compared to 28 percent for WIC, see Chapter 3). The greater understanding of offset rules among the EWIC subjects supports the conclusion that more extensive outreach and communication regarding the offset can improve understanding, but not by much, even among those motivated to volunteer.

We draw two implications from the Stage 2 results. First, it is possible to increase understanding of the offset rules through additional outreach to beneficiaries. This is seen in comparing the T21 group to the T1 group and in the direct evidence comparing the T22 group to the T21 group. In the next section, we suggest that information provided under a national offset program would be greater than the limited outreach to T1 subjects and that this would likely lead to somewhat greater understanding of the offset policy.

Second, while it is reasonable to think that greater understanding of the offset policy's incentive should lead to a larger behavioral response to the offset, the demonstration provides no evidence of a larger behavioral response. Although we find that the T22 group had a higher level of correct understanding than the T21 group, the T22 group does not have higher levels of employment or the percentage with earnings above BYA. Thus, even though we expect a permanent offset program would result in better understanding than that observed in the T1 group, the lack of positive evidence for the connection between understanding of rules and behavioral response dampens our expectation of larger-than-Stage-1 behavioral impacts for a permanent program.

Limitations of Implementation

In order for estimates from the BOND evaluation to match the would-be impacts of a national offset policy, demonstration implementation must simulate the conditions that would exist under a national policy. Ideally, in order to simulate conditions that would exist under a national policy, outreach and information provided to beneficiaries and their trusted advisors in Stage 1 would be as robust as they would be under a national program. Similarly beneficiaries would have as much confidence that SSA will adjust their benefits according to the offset as they would under a national program.

The outreach to Stage 1 treatment subjects was designed to inform them about the nature of the offset and their opportunity. However, that outreach was more limited than what would likely occur in a national benefit offset program. The outreach consisted of only two letters from the implementation team, two phone call attempts, one notice from SSA, and the provision of information to other local stakeholders likely to be the trusted advisors of at least some beneficiaries. By contrast, it seems likely that under a national policy, SSA would conduct additional steps to inform beneficiaries in addition to those taken for the demonstration. For example, local field offices would likely provide written information (e.g., brochures and fact sheets). In addition, SSA would update its Red Book – *A Guide to Work Incentives* and website to explain the rules. We also assume that SSA would revise operating procedures and staff

training to explain the new rules. All of SSA's notices and other communications with SSDI beneficiaries would also contain description of the benefit rules.

In BOND, informing stakeholders that a small share of beneficiaries residing in the site jurisdictions had a legitimate opportunity to use the offset was inherently challenging. Certainly such minimal efforts would have been insufficient to provide stakeholders with a robust understanding of the offset. Consistent with this concern, the BOND process analysis found anecdotal reports of VR counselors, SSA field office staff and other presumably trusted professionals providing incorrect information to treatment subjects early in the demonstration. Those reports included some stakeholders advising beneficiaries that BOND was a scam. We have no way of knowing how commonly treatment subjects encountered such misinformation, however, or how they responded.

In addition to stakeholder understanding, two other aspects of implementation suggest concern about how well demonstration results simulate a national policy change. First, many offset users encountered long delays before SSA first adjusted their benefits. Controls also experienced delays, but as was discussed in Chapter 4, work CDR processing times (one of the main factors that contribute to benefit adjustment delays) were shorter for control subjects than for treatment subjects. If not understood, delays in benefit adjustments may confuse beneficiaries, rather than either reinforce understanding or correct any misunderstanding. Delays may also undermine beneficiary trust that SSA will follow the rules, rather than strengthen it.

The final aspect of implementation that may have affected behavioral responses is the five-year time limit on the BOND participation period (BPP). One reason why the BOND time limit might limit the earnings of treatment subjects is discussed in Bell et al. (2011). Treatment subjects might be concerned about continuation of SSDI benefit eligibility after the BPP ends and whether work during the BPP would affect SSDI entitlement after the BPP ends. If so, they might stop earning above BYA before the end of the participation period to demonstrate their inability to engage in SGA after the BPP ends; or they might not earn above BYA at all. However, careful study of the BOND evidence does not suggest that longer or unlimited duration would have affected impacts during the first five years (i.e., there is no indication of impacts decreasing in the later years of the follow-up period). Past large scale studies provide mixed evidence about whether the length of an intervention matters. The Negative Income Tax experiments conducted in the 1960s and 1970s found that increasing duration from three years to five resulted in larger impacts in the first three years (Robins and West 1980). On the other hand, the National Health Insurance Experiment conducted in the 1970s and 1980s tested both 3 and 5 year interventions and found that the length of intervention had no impact on findings (Newhouse 1993).

Summary
This review of reasons why more BOND treatment subjects did not use the benefit offset suggests that several factors may contribute: (i) beneficiary capacity to engage in SGA, (ii) the implicit marginal tax rate implied by the $1 for $2 benefit offset, (iii) the inherent complexity of the earnings rules, (iv) delays in benefit adjustment which may have led to confusion and undermined trust in the rules, (v) the five-year time limit on offset use, and (vi) insufficient outreach to subjects to inform them about the offset. The first three of these factors would be present under a permanent, national offset program with the same rules as the BOND benefit offset. The last three factors might have contributed to lower offset use in the demonstration than would be seen in a permanent, national offset program. Benefit adjustment delays and the five-year limit possibly affected both stages of BOND, but the last factor would possibly have been less salient for Stage 2 (because of informed consent) and even less salient for T22s (because of EWIC).

Given the more robust outreach and better understanding observed in Stage 2,the minimal improvement in understanding produced by the resource-intensive EWIC services, and the fact that these were volunteers and therefore the most likely to use the offset, it seems likely that the impact on the proportion earning above BYA in the Stage 2 sample (which at 23 percent is more than three times greater than the 7 percent relative effect found in Stage 1) is an upper limit impact of a national offset policy. The impact on proportion earning above BYA might have been somewhat larger in the nationally-representative Stage 1 had outreach to treatment subjects been more robust, benefit adjustments been made in a more timely manner, and if the offset were not time-limited. However, there is nothing in the evidence to suggest that the magnitude of the effect under an established national program would reach the effect found for volunteers in Stage 2. Further, the results from Stage 2 imply that even with a 23 percent effect, the offset policy would still increase SSDI benefits due to beneficiaries.

7.3. Enhancements to Work Incentives Counseling Did Not Advance BOND's Goals

Stage 2 of BOND tests whether enhancements to standard work incentives counseling services improves beneficiary employment outcomes when a benefit offset incentive is in place. This section considers the success of that test and the lessons it offers for the value of more extensive counseling. In short, the enhancements tested did not yield greater employment and earnings under the BOND offset. This is true even though per-beneficiary counseling expenses for EWIC were more than twice those for WIC.[130]

Evidence in Chapter 3 shows that subjects assigned to receive EWIC had a much higher contact rate with counselors and were much more likely to receive services beyond information and referral (I&R) than those randomized to WIC. Moreover, EWIC counselors' use of EWIC-specific counseling tools met or came close to the benchmark rates set in the design of BOND. Finally, random assignment to EWIC resulted in a greater share of beneficiaries assigning a ticket under Ticket to Work (26 percent, compared to 19 percent for those served by WIC counselors) and enrolling in state VR programs (15 percent, compared to 12 percent).

Evidence from the process analysis clearly shows that BOND was successful in testing the EWIC model. Relative to the WIPA counseling model, the EWIC enhancements yielded positive effects on some outcomes: 1) at least some improvements in beneficiaries' understanding of the benefit offset rules; 2) shorter average duration from first offset use to benefit adjustment; and 3) lower average overpayments. However, counseling enhancements did not: 1) increase use of the offset; 2) generate higher earnings; or 3) reduce SSDI benefits. Nor did the evaluation find any evidence that the enhancements improved beneficiaries' lives in other areas such as health status, health insurance coverage, participation in other income assistance programs, or household income.

7.3.1. Why Enhanced Counseling Had No Payoff

The findings raise the question: Why did EWIC have no impacts on earnings or benefits? EWIC services might increase employment and earnings more than WIC services for two reasons. First, EWIC staff might help beneficiaries better understand the offset's relatively advantageous consideration of earnings.

---

[130] On a per-beneficiary basis, counseling expenditures for the average EWIC subject were 2.14 times greater than for the average WIC subject.

The survey results reported in Chapter 3 indicate EWIC subjects' understanding of the offset is only slightly better than WIC subjects' understanding: 52 percent of the T22 group and 48 percent of the T21 group accurately stated that, under the offset, higher earnings reduce benefits but not to $0 (the difference is statistically significant). This small difference may understate differences in understanding, because enhanced counseling may have deepened the understanding of those who already had a basic understanding of the offset, rather than result in a shift from a basically incorrect understanding to a basically correct one.

The low absolute rate of correct understanding for EWIC subjects, despite the considerable enhancements intended to help them grasp the benefit implications of earnings under the BOND offset, suggests that the ability of beneficiaries to understand complex earnings rules limits the ability of rule changes to affect beneficiary work behavior. EWIC made a concerted effort to reach all T22 beneficiaries and a high proportion of beneficiaries assigned to EWIC (95 percent) received counseling services. The high take up, coupled with the monthly contact associated with EWIC counseling suggests that the understanding of the T22 subjects may be as high as could reasonably be achieved among similar beneficiaries under an ongoing national benefit offset policy.

The greater use of employment services provided by TTW among T22 subjects suggests another route to higher earnings for T22 subjects: increased work readiness and productivity. This route requires that services lead to beneficiaries being substantially more prepared for employment or that services raise beneficiaries' earnings potential. The research literature casts doubt on whether they do. A large body of evidence shows that SSDI beneficiaries assigning their tickets under TTW have higher employment and earnings than beneficiaries who are comparable based on the limited characteristics available from administrative data. Such studies are unable to adjust for many unobserved differences—differences that likely had an important influence on the decision to use TTW. To date, no study of VR agencies or other ENs has definitively shown that VR services have positive impacts on employment outcomes.[131]

---

[131]   Existing studies are not definitive because while they find positive differences between participant outcomes and outcomes of nonparticipants, the studies have been unable to adequately control for confounding factors—most notably, self-selection of participants. Dean et al. (1999) used quasi-experimental methods that received a moderate rating from the Department of Labor's Clearinghouse for Labor Evaluation and Research and found positive impacts of VR services during the pre-TTW period. However, that paper did not differentiate between SSDI beneficiaries and other VR clients.

   More recently, Dean et al. (2015) used a different quasi-experimental approach and finds positive impacts on employment and earnings for some categories of services provided to Virginia VR clients with mental illnesses, but do not differentiate between SSDI beneficiaries and others.

   Stapleton et al. (2014) used the randomized mailing of tickets to beneficiaries at the start of TTW to measure the incremental impacts of TTW on service use by SSDI beneficiaries and later employment and benefit outcomes. That paper found significant impacts on service use, but no impacts on earnings or months of SSDI non-payment status following suspense or termination for work (NSTW months).

   Schimmel, Hyde and Stapleton (2015) found considerable expansion of TTW use following regulatory changes in 2008. That paper also found that TTW participants continue to be much more likely than other beneficiaries to have NSTW months. However, that paper could not differentiate between the impact of the post-2008 TTW expansion on NSTW months and confounding factors, including self-selection and the effects of the 2007-2009 recession.

Moreover, the 3-percentage impact on receipt of such services observed for Stage 2 treatment subjects would need to produce very large earnings impacts in order for *average* earnings across *all* subjects to rise to a substantively and statistically important degree.[132]

7.3.2. Implications for Future Counselor-Assistance Initiatives

Enhancements to work incentives counseling moved a small share of beneficiaries into a modestly more accurate understanding of the offset's work incentives and a modest increase in use of TTW and state VR services. However, there were no earnings gain.

These findings imply that enhancements to work incentive counseling that involve more proactive, frequent contact with beneficiaries and more frequent provision of work incentives counseling beyond I&R are of themselves very unlikely to lead to higher earnings. Those specific elements of counseling are the most distinguishing features of the enhancements. They were 3 to 10 times more prevalent for EWIC recipients than for WIC recipients (depending on the element and year).

7.4. Lessons for Research on the SSDI Earnings Rules

This final section uses the BOND results to inform future tests of changes in the SSDI earnings rules in three areas: (a) the limitations of using informed volunteers; (b) the interpretation of evidence when there are long delays in the benefit adjustment process; and (c) types of changes to the earnings rules that might be worth testing in the future.

7.4.1. Limitations of Using Informed Volunteers to Test SSDI Work Incentives

Broadly speaking there are two strategies for testing innovation in ongoing programs. One approach randomizes everyone (meeting some set of criteria); the other approach randomizes volunteers (but not non-volunteers). Because BOND used both strategies to test the offset plus standard work incentives counseling (WIC)—everyone in Stage 1, only volunteers in Stage 2—it provides a rare opportunity to consider the relative merits and limitations of testing the two strategies.

Use of informed volunteers in a social experiment is most likely to work well if the volunteers include the bulk of those for whom the intervention is likely to affect behavior. Because the BOND offset clearly provides a very favorable treatment of annual earnings above BYA relative to current law, we expected the recruitment effort to attract the bulk of those for whom the BOND offset would be of use. If, in the extreme, all subjects in the Stage 2 recruitment pool who would have used the offset during the demonstration period had they volunteered and been assigned to a treatment group, then the impacts for the Stage 2 volunteers would have constituted 100 percent of the impacts for all SSDI-only beneficiaries.

BOND is unusual in that Stage 1 provides the opportunity to assess the extent to which this expectation was realized. Comparison of the findings for the Stage 2 T21 subjects to those for the Stage 1 SSDI-only

---

[132]　As an example, consider a $15,817 impact on 2012-2015 earnings—an effect equal to half the difference in earnings between C2 subjects employed at baseline and those not employed. This average gain in the "swing group" translates into a $989 average impact for the T22 sample as a whole. A measured impact of $989 would not be large enough to be statistically significant and would add little in substantive terms (just $247 per year) to total earnings, making detection of a true average impact of this size—should one occur—unlikely. In all likelihood, the impacts of employment services are much smaller than this illustration.

T1 subjects[133] indicates that Stage 2 only recruited about 1 out of every 5 subjects in the Stage 2 recruitment pool who would have used the offset over the next five years had they volunteered and been assigned to T21. Specifically, we estimate that 21 percent of SSDI-only offset users in the T1 group (95 percent confidence interval: 18.4 percent to 24.4 percent) would have volunteered for Stage 1.[134] Thus, had BOND used only the Stage 2 volunteers, on the order of 4 out of 5 of potential offset users in the SSDI-only population would not have been represented in the demonstration at all.[135]

Basing inferences on the Stage 2 volunteers alone would not only be problematic because it would get the number of beneficiaries who would use the offset wrong, but also because it would misrepresent patterns of offset usage. Consider those with earnings above BYA. Compared to T21 beneficiaries (*the cases the volunteer-only study would have to rely on*), T1 SSDI-only subjects who would not have volunteered (*the cases the volunteer-only study would exclude*) were less likely to have increased their earnings to above the BYA amount and were more likely to have received a windfall. That is, the point estimates imply that for every month in which T21 subjects had earnings above BYA as a result of the offset, SSA paid about 2 months of at least partial benefits to those who received windfall benefits because of the offset. For those T1 SSDI-only subjects who would not have volunteered, for every month where earnings would have been above BYA, SSA would pay 7 months of windfall benefits.[136]

---

[133] We focus on SSDI-only T1 subjects because SSDI/SSI concurrent subjects were not in the Stage 2 solicitation pool.

[134] Under the assumption that everyone who would have used the offset volunteered, we have two estimate of the rate of offset use:
   (i) the simple estimate among the T1 subjects—i.e., the fraction of the T1 subjects using the offset in the sample of the entire population of SSDI-only beneficiaries.
   (ii) the product of the fraction of T21 subjects using the offset and the fraction of beneficiaries volunteering. In fact, the estimate among the T1 subjects is 3.80 percent, while the estimate from the T21 subjects is only 0.81 percent (15.26 percent of T21s multiplied by the 5.321 percent volunteer rate). Put differently, the T1 rate is about 5 times the implied T21 rate. See Appendix Exhibit H-2.

[135] These results raise the question of why some beneficiaries who would have gained from the benefit offset did not volunteer for Stage 2 when solicited. Several reasons are possible, including lack of understanding of the demonstration, mistrust of the demonstration staff making the offer, low expected value of volunteering because of the uncertain assignment to the treatment groups, and poor health at the time of the offer that later improved unexpectedly.

[136] These ratios were calculated from the point estimates for T21 impacts and inferred impacts for T1 SSDI-only subjects who would not have volunteered for the 2012 through 2015 period. Specifically, we: 1) divided the point estimates for impact on years with earnings above BYA (0.08 for T21 and .0074 for the SSDI-only T1 subjects who would not have volunteered) by 4 to estimate the percentage of months in years in which earnings above BYA due to the offset (2.0 percent and 0.19 percent, respectively); 2) divided the estimates on months with positive benefits due (2.13 for T21 and 0.63 for the SSDI-only T1 subjects who would not have volunteered) by 48 to determine the impact on the percentage of months with positive benefits due (4.4 percent and 1.3 percent, respectively), and 3) divided the latter by the former to obtain the reported ratios, respectively. Appendix H compares inferred estimates of impacts of a larger set of earnings and benefit outcomes for all SSDI-only T1 subjects who would not have volunteered for Stage 2 to the T21 impact estimates, each for the 2012—2015 period.

There are, of course, good reasons to use informed volunteers. Perhaps most importantly, testing policies with informed volunteers helps ensure that participation is in the best interest of the subjects. SSA's current demonstration authority, as established under the 2015 Bipartisan Budget Act, requires SSA to use volunteers.[137] To our knowledge, prior to BOND, all of SSA's tests of work incentives that have randomly assigned individuals to treatment or control groups have used informed volunteers exclusively.[138] For situations in which there is no practical alternative to use of volunteers, the BOND findings point to the importance of designing the demonstration: 1) to maximize enrollment by beneficiaries likely to be affected by the innovation; and 2) to test features that would help evaluators predict how the innovation would affect the behavior of non-volunteers.[139]

7.4.2.   Sensitivity of Findings to Implementation Conditions

The BOND results only directly generalize to a national offset implemented under conditions similar to BOND—in particular, beneficiary understanding of the offset, the speed with which SSA adjusted benefits, and the frequency and size of overpayments caused by any delays. If implementation conditions under a national policy are different from those established in the demonstration, then BOND Stage 1 estimates may not extrapolate to a national offset policy.

It is certainly possible that implementation conditions under a national policy would be different than they were for treatment subjects. As pointed out in Section 7.3.2, comparison of the 29 percent of T1 subjects with a basic understanding of the effects of sustained, substantial earnings on benefits under the offset to the 54 percent of C1 subjects with a basic understanding of effects under current law—the current national policy—suggests that beneficiary understanding of the offset under a national policy might be better than that of T1 subjects.[140] In addition, SSA is already attempting to reduce adjustment delays under current law in important ways.[141] Further, early technical issues that delayed automated adjustments, but were subsequently corrected, would presumably not occur under a national policy, and SSA could potentially address technical issues that remained at the end of the evaluation period (see Chapter 4).

---

[137]   SSA's demonstration authority is described in Section 234 of Title II of the Social Security Act. When BOND was initiated, SSA's demonstration authority did not require the use of volunteers in demonstrations.

[138]   Examples predating BOND include Project NetWork, the State Partnership Initiative states that used random assignment, the Youth Transition Demonstration, the Mental Health Treatment Study, Accelerated Benefits, and the Supported Employment Demonstration. The current Promoting Opportunity Demonstration (POD) also uses volunteers, as mandated by SSA's current demonstration authority.

[139]   Examples include collecting more detailed baseline information about volunteers and non-volunteers, and building in randomized outreach experiments, to assess the sensitivity of the findings to exogenous changes in the volunteer rate.

[140]   The BOND findings suggest that implementation of a national offset policy would lead to better understanding. Unlike a demonstration, in a national policy stakeholders would have better understanding of the rules and beneficiaries would learn from each other.

[141]   SSA's current efforts are attempting to address both reporting delays (e.g., providing a means for beneficiaries to report their earnings via the internet and developing the ability to scan large private payroll databases on a monthly basis) and processing delays (e.g., by use of data analytics to prioritize work CDRs and devoting more resources to this task).

The BOND findings do not provide direct evidence of how improvements in beneficiary understanding relative to the understanding of T1 subjects, or reductions in adjustment delays, would affect earnings and benefit outcomes in a national policy. Given the more robust outreach and better understanding observed in Stage 2, and the minimal improvement in understanding produced by the resource-intensive EWIC services, it seems likely that the impact on the proportion earning above BYA in the Stage 2 sample (which at 23 percent is more than three times greater than the 7 percent relative effect found in Stage 1) is an upper limit impact of a national offset policy. It thus seems plausible to infer that the impact on the proportion earning above BYA might have been somewhat larger in the nationally-representative Stage 1 had outreach to treatment subjects been more robust, benefit adjustments been made in a more timely manner, and if the offset were not time-limited. However, there is nothing in the evidence to suggest that the magnitude of the effect under an established national program would reach the effect found for volunteers in Stage 2. Further, the results from Stage 2 imply that even if the benefit offset increases the share earning more than BYA (relative to controls) by 23 percent, the benefit offset would still increase SSDI benefits.

These are important sources of uncertainty about the implications of BOND's findings for a national program. Their existence and potential importance suggest that the future tests of changes to the earnings rules would benefit from including design features that allow the evaluators to assess the sensitivity of the findings to efforts to improve beneficiary knowledge, adjustment delays, and any other implementation conditions of potentially critical importance to the interpretation of the findings.[142] The information generated would help SSA understand the benefits and costs of efforts to improve beneficiary knowledge about the earnings rules, reduce adjustment delays, or otherwise improve the delivery of services to beneficiaries.

7.4.3.   Testing Other Designs for the SSDI Earnings Rules

Because of BOND, it is now clear that adopting a $1 for $2 benefit offset as implemented in BOND as a national policy would neither substantially increase beneficiary earnings, nor reduce the liabilities of the Disability Insurance Trust Fund.[143] Furthermore, while plausible improvements to speed of adjustment and understanding might moderately increase earnings, the BOND results imply that it is not reasonable to expect earnings increases to be sufficient to reduce the Trust Fund's net liabilities.

Policymakers have already directed SSA to evaluate the impacts of a set of changes that, relative to the BOND offset, puts more emphasis on the goal of reducing Trust Fund liabilities—namely the rules to be

---

[142]   An attractive option is to conduct, within the demonstration, randomized tests of efforts to: 1) increase beneficiary knowledge (for example, varying levels of outreach, or how information is presented); 2) expedite benefit adjustments; or 3) change other potentially important conditions of implementation. Such activities could be tested within each treatment or control group in a manner that does not necessarily require larger sample sizes.

[143]   Consistent with the BOND evaluation analysis plan and the results presented in this report, the statement ignores any increase in entry induced by the offset. (See Maestas, Mullen, and Zamarro 2010 for a summary of research on induced entry in response to an SSDI benefit offset.) Induced entry into the SSDI program from such post-entitlement policy changes, if any should occur, would further worsen the implications for the Trust Fund.

tested in the Promoting Opportunity Demonstration (POD).[144] Some policymakers also have expressed interest in designs that would put more emphasis on the goal of increased earnings. The balance of this section considers BOND's implications for both sets of changes.

Designs to Improve the Financial Status of the DI Trust Fund

We begin by considering POD. Three features of POD are intended to reduce benefits due to SSDI beneficiaries and thus net DI Trust Fund liabilities relative to current law:

1. Like BOND, POD has a $1 for $2 offset that provides higher incomes for those earning more than the SGA amount. However, while BOND's offset begins at the SGA amount, POD's offset begins at about three-quarters of the SGA amount. Thus, relative to BOND and current law, benefits would be lower for those with earnings in this under-SGA range. In addition, because benefits would be lower under POD, those with earnings in this range income would be lower than under either BOND or current law.

2. The POD rules eliminate TWP and GP months. Instead, for a given level of earnings during what under current law would be TWP and GP months, the POD offset reduces benefits and thus total beneficiary income.

3. The POD rules substantially limit the extent to which reporting IRWE reduce earnings count for benefit adjustment purposes. Current law reduces countable earnings dollar for dollar with allowable IRWE. The POD offset treats the TWP limit as a minimum IRWE amount for all beneficiaries. As a result only allowable IRWE in excess of the TWP limit can be used to reduce countable earnings, so for given levels of earnings and IRWE, SSDI benefits will be lower. Relative to current law, this provision of POD will reduce the marginal value of earning more than the SGA amount when monthly earnings net of IRWE are less than the SGA amount.

The BOND findings have several implications for the potential behavioral effects of a POD-like offset policy.[145] First, the BOND results suggest that the increase under POD in the proportion of beneficiaries earning more than the BYA will be smaller than the five year 0.4 percentage point increase for BOND Stage 1 treatment subjects (Exhibit 5-2). This is because, compared to the BOND rules, the POD offset pays less benefits for a given level of earnings.

Second, POD reduces benefits for those with earnings slightly less than SGA. This change might induce some beneficiaries with earnings in this range to earn more. However, the BOND offset induced only a small share of those who otherwise would have had earnings in this range to earn more than the annual

---

[144] See SSA (n.d., accessed November 16, 2017) for additional information about POD.

[145] This discussion concerns a national policy, and ignores implementation conditions that are specific to the POD demonstration: namely the use of volunteers only and the fact that volunteers assigned to the POD design may choose to withdraw from the study and return to current law at any time as required under current SSA authority provided in section 234 of the Social Security Act. These provisions make it much less likely that POD volunteers will include those whose incomes are most adversely affected by the POD design, and that those volunteers assigned to treatment will behave as they would under a national policy should they decide that the POD design is adversely affecting their incomes relative to current law. These circumstances make it more likely that evaluators will not observe in POD the earnings behavior that would occur under a permanent, national POD-like offset policy.

equivalent of the SGA amount (Exhibits 7-1 and 7-2). It is possible that, because the loss of benefits increases the value of additional earnings to the beneficiary, a POD-like offset would induce more beneficiaries who otherwise would have earnings in this range to increase their earnings. However, the BOND finding may mean that many are either unable to do so, or found that the loss of $1 in benefits for every additional $2 in earnings made additional earnings unattractive. In fact, a POD-like $1 for $2 offset might encourage some whose earnings would be between the TWP limit and the SGA amount under current law to *reduce* their earnings to the TWP limit.

Third, the BOND findings suggest that the POD offset by itself—in the absence of change in the treatment of IRWE—would lessen use of IRWE. Consistent with theory, we found that the BOND offset reduced use of IRWE. As a result, the presumably negative impact on IRWE of adopting POD-like IRWE rules along with the POD offset might be substantially less than the effect of adopting the same IRWE rules without the POD offset.

Designs to Increase Beneficiary Use of Earnings Capacity
The second potential direction for future research could focus on the second goal of BOND, increasing beneficiary earnings and, consequently, income. Future research of this sort would help policymakers understand the economic costs of rules that discourage work and thus to move towards a policy that maximizes earnings and income for a given level of benefits (Fichtner and Seligman 2016). Such evidence would also help policymakers understand the implications of proposals to move away from SSDI's "inability to work" definition of eligibility toward a definition that is based on residual work capacity (Social Security Advisory Board 2006) or on a significant medically determinable impediment to work (Jacobson et al. 2015).

On the one hand, the BOND findings could be interpreted as implying this line of research is of little value. The lack of a large impact on the proportion with earnings above BYA under BOND is consistent with the hypothesis that very few beneficiaries who do not engage in SGA under current law have the capacity to do so. On the other hand, as discussed in Section 7.2, more beneficiaries may have the capacity to engage in SGA than revealed by BOND and might do so if their income rose faster with earnings.

To distinguish between these alternative interpretations of the BOND results, one might design a demonstration in which treatment subjects receive full benefits no matter how much they earn. Such a rule might be problematic as a national policy, because of the likely large net cost to the DI Trust Fund. Even so, a test would reveal how much beneficiaries can earn and would therefore provide a stronger foundation for developing more efficient earnings rules than the findings from either current law or BOND. That is because we would know how much beneficiaries can earn when earnings rules do not discourage work. In contrast, under both current law and the BOND offset, we do not know the extent to which earnings are limited because of earnings rules versus medical conditions and other factors.

To illustrate how the findings of such a demonstration could be used to examine a specific policy, consider a policy that would pay some groups a monthly amount that is less than 100 percent of the current law benefit (possibly after a time-limited period of initial eligibility), but continue to pay them the same amount regardless of how their earnings change thereafter. The demonstration data could be used to identify groups that could be targeted for such reductions, based on their characteristics and their behavior in the absence of a reduction. For instance, groups that attained much higher levels of earnings in the demonstration than under current law could be offered a partial benefit under a new policy, not

conditioned on earnings. Holding earnings constant at the demonstration level, such a design would preserve the rate at which earnings are converted to net income at the margin and reduce SSDI benefits due to beneficiaries. Further, the reduction in benefits would likely induce them to earn more to make up for the lost income.

A disability wage tax credit (DWTC), which has received greater attention among policy analysts, provides a more complex example (Burkhauser and Daly 1996; Mashaw and Reno 1996; Social Security Advisory Board 2006; Gokhale 2013; Stapleton and Hyde 2017). A DWTC is similar to BOND in that it changes the marginal value of an additional dollar of earnings to the beneficiary. Whereas the BOND $1 for $2 offset imposes an implicit 50 percent marginal tax on countable earnings above BYA, however, a DWTC would include a marginal earnings subsidy (negative marginal tax) from zero earnings up to a specified earnings threshold, no marginal tax or subsidy above the first threshold up to a second threshold, then a marginal tax for earnings above the second threshold until the value of the subsidy reaches zero. Demonstration and current law data could be used to design a DWTC that, holding earnings constant, results in lower transfer payments to beneficiaries who earn much more under demonstration rules than under current law, but to the extent feasible, maintains earnings incentives that are close to those in the demonstration.

8.   Summary

This chapter summarizes the main findings from the BOND evaluation's process, participation, impact, and cost-benefit analyses for this *Final Evaluation Report*.

8.1.   Origins of the $1 for $2 Benefit Offset Demonstration

The BOND evaluation satisfies the requirement of The Ticket to Work and Work Incentives Improvement Act of 1999.[146] Since the 1990s, rising program costs led policy makers to consider changes to SSDI program rules that might increase work and earnings and reduce total SSDI benefits. The total loss of SSDI benefits after a sustained period of substantial earnings, the "cash cliff" in current law, is widely presumed to discourage work. The legislation required SSA to conduct an evaluation to produce nationally representative estimates of the costs and benefits of an ongoing national $1 for $2 benefit offset for the SSDI program—replacing the "cash cliff" with a ramp.

While the Ticket to Work Act directed SSA to identify ***reductions*** in federal expenditures that may result from the permanent implementation of a $1 for $2 benefit offset, previous analysis by SSA actuaries suggested that such a policy might ***increase*** SSDI benefits costs, for two reasons. First, a benefit offset would pay partial benefits in months in which earnings would result in no benefits under current law. Second, a benefit offset policy could induce individuals to enter the SSDI program who would not do so under current law (Mashaw and Reno 1996; Tuma 2001). However, after considerable consultation with experts, SSA determined that it would not be feasible for BOND to measure induced entry using an experimental research design.

From 2004 to 2010, SSA conducted the Four-State Benefit Offset Pilot Demonstration, a random assignment study that tested a $1 for $2 benefit offset on earnings above the SGA amount with a sample of beneficiary volunteers. The four-state pilot found that relative to current law, a benefit offset increased the proportion of beneficiaries with earnings above an annual equivalent of SGA. However, the pilot found no effect on average earnings and found that SSDI benefits increased (Weathers and Hemmeter 2011). Finally, the pilot developed recommendations for ways that SSA could improve administrative procedures for the national test of the benefit offset policy.

8.2.   Hypothesized Effects of a $1 for $2 Benefit Offset

For the benefit offset tests, the theoretical direction of impacts on mean earnings and mean benefits is ambiguous.[147] This ambiguity arises because the incentives created by the benefit offset vary with what a

---

[146]   Ticket to Work and Work Incentives Improvement Act of 1999, Public Law 106-170, Section 302, 106th Congress, codified at U.S. Code 42 (1999) §434.

[147]   Total earnings and total SSDI benefits are the evaluation's two predesignated confirmatory outcomes. These outcomes address the primary policy objectives for the demonstration's benefit offset policy. We apply a higher standard of evidence to the confirmatory outcomes by adjusting the p-values of the impacts estimates to account for multiple comparisons. This adjustment ensures that the possibility of one or more "false positive" findings of impact on the confirmatory outcomes is kept to 10 percent within each of three sets of tests (Stage 1, Stage 2 tests of the benefit offset versus current law, and State 2 tests of enhanced counseling versus standard counseling).

beneficiary's earnings would be under current law. Opposite effects are expected for those who would have had earnings below BYA under current law (hereafter the "below BYA group") and those who would have had earnings above BYA under current law (hereafter the "above BYA group").

- The offset is expected to induce treatment group subjects in the below BYA group to have on average (i) higher earnings and (ii) lower SSDI benefits than they would under current law. The lower SSDI benefits result from the change in earnings behavior by those induced by the offset to increase earnings above the BYA threshold. Those who are induced to increase their earnings above BYA receive partial SSDI benefits under the offset policy rather than full benefits— leading to lower benefits for this group.

- Conversely, the offset is expected to cause treatment subjects in the above BYA group to have on average (i) lower earnings (though still above BYA) and (ii) higher SSDI benefits than they would under current law. These higher SSDI benefits result from two causes. First, there is a mechanical effect, by which the offset rules will provide partial benefits in months when current law would have suspended or terminated benefits. This provides most of those in the above BYA group with a windfall for no change in behavior.[148] Second, the increase in income from the partial benefits will induce some to lower their earnings. This change in earnings behavior serves to further increase SSDI benefits.[149]

The net impacts on earnings and benefits combine effects on these two types of subjects.

- For the impact on total earnings to be positive, the *positive impact* on earnings expected for the below BYA group would have to be *larger* than the *negative impact* on earnings expected for the above BYA group.

- Similarly, for the impact on SSDI benefits to be negative, the *reduction in benefits* for the below BYA group must be *larger* than the *increase in benefits* for the above BYA group.

If the benefit offset has no or little effect on earnings behavior, then the non-behavioral windfall effect will dominate, causing an increase in total SSDI benefits.

8.3.   The BOND Evaluation

SSA implemented BOND to address the requirements in the Ticket to Work Act. SSA conducted BOND in 10 of SSA's 53 area offices, selected at random to represent the nation. The BOND evaluation used random assignment to test the $1 for $2 benefit offset. BOND includes two stages.

- *Stage 1* tests how a national benefit offset would affect earnings and program outcomes for the entire SSDI population. In this stage, the demonstration randomly assigned beneficiaries into either a treatment group "T1" (subject to benefit offset rules and offered Work Incentives Counseling [WIC]) or a current-law control group "C1".

---

[148]   High-earning beneficiaries whose earnings above BYA are more than two times their benefit amount will be in "full offset" and so will not be due any SSDI benefits.

[149]   The reduction in earnings by some above BYA group beneficiaries reduces the size of the benefit offset.

- *Stage 2* tests the impact of the offset for those expected to be most likely to use the offset—recruited and informed volunteers. Stage 2 also tests the extent to which enhanced counseling (EWIC) affects impacts. In this stage, the demonstration randomly assigned volunteers into one of three assignment groups: a treatment group "T21" (benefit offset rules and offered WIC), a second treatment group "T22" (benefit offset rules and offered EWIC), or a current-law control group "C2".

BOND provides three tests of the benefit offset policy versus current law:

- Stage 1's T1 versus C1 comparison;

- Stage 2's T21 versus C2 comparison; and

- Stage 2's T22 versus C2 comparison.

Stage 2 of the demonstration provides the additional test of enhanced work incentives counseling (EWIC) versus standard work incentives counseling (WIC) for beneficiaries subject to the benefit offset rules:

- Stage 2's T22 versus T21 comparison.

8.4. Findings from the BOND Evaluation

Exhibit 8-1 restates the key results from the BOND evaluation. Each panel of the exhibit provides the main findings from the impact, participation, process, or benefit/cost analyses. The results for the Stage 1 treatment group appear in the first column, and those for the two Stage 2 treatment groups appear in the second and third columns.

Exhibit 8-1.   Major Evaluation Findings, by Study Component and Intervention

	Target Population and Intervention		
	Stage 1 Full SSDI Caseload	Stage 2 SSDI-Only Volunteers	
Indicator	The Benefit Offset (T1)	The Benefit Offset (T21)	The Benefit Offset + EWIC (T22)[a]
Impact Analysis			
	Relative to current law, the benefit offset…	*Relative to current law, the benefit offset…*	*Relative to current law, the benefit offset plus EWIC…*
Total Earnings	had no detectable effect on total earnings during 2011–2015. Confidence interval[b] of -3 to +3 percent of C1 mean.	had no detectable effect on total earnings during 2012–2015. Confidence interval[b] of 0 to +18 percent of C2 mean.	had no detectable effect on total earnings during 2012–2015. Confidence interval[b] of -5 to +25 percent of C2 mean.
SSDI benefits	increased SSDI benefits $665 per beneficiary during May 2011–Dec. 2015 ($143 per year, or $12 per month, or 1.2 percent of C1 mean). Confidence interval[b] of 0.9 to 1.6 percent of C1 mean.	increased SSDI benefits $1,791 per beneficiary during 2012–2015 ($448 per year, or $37 per month, or 3.6 percent of C2 mean). Confidence interval[b] of 1.5 to 5.7 percent of C2 mean.	increased SSDI benefits $1,997 per beneficiary during 2012–2015 ($499 per year, or $42 per month, or 4.0 percent of C2 mean). Confidence interval[b] of 1.6 to 6.4 percent of C2 mean.
Earnings above BOND Yearly Amount (BYA)	increased the proportion of beneficiaries with earnings above BYA during 2011–2015 by 0.4 percentage points (7 percent of C1 mean of 5.1 percent). Confidence interval[b] of 3 to 12 percent of C1 mean.	increased the proportion of beneficiaries with earnings above BYA during 2012–2015 by 4.0 percentage points (25 percent of C2 mean of 16.0 percent). Confidence interval[b] of 12 to 38 percent of C2 mean.	increased the proportion of beneficiaries with earnings above BYA during 2012–2015 by 3.3 percentage points (21 percent of C2 mean of 16.0 percent). Confidence interval[b] of 8 to 33 percent of C2 mean.

	Target Population and Intervention		
	Stage 1 **Full SSDI Caseload**	**Stage 2** **SSDI-Only Volunteers**	
Indicator	**The Benefit Offset (T1)**	**The Benefit Offset (T21)**	**The Benefit Offset + EWIC (T22)[a]**
Number of months with SSDI benefits	increased the number of months with SSDI benefits by 0.8 months during May 2011–Dec. 2015 (1.6 percent of C1 mean). Confidence interval[b] of 0.9 to 1.6 percent of C1 mean.	increased the number of months with SSDI benefits by 2.1 months during 2012–2015 (5.0 percent of C2 mean). Confidence interval[b] of 3.7 to 6.4 percent of C2 mean.	increased the number of months with SSDI benefits by 2.1 months during 2012–2015 (4.9 percent of C2 mean). Confidence interval[b] of 3.4 to 6.4 percent of C2 mean.
Any report of non-countable earnings	decreased the proportion reporting non-countable earnings by 0.29 percentage points during May 2011–Dec. 2015 (36 percent of C1 mean). Confidence interval[b] of -47 to -25 percent of C1 mean.	decreased the proportion reporting non-countable earnings by 1.18 percentage points during 2012–2015 (30 percent of C2 mean. Confidence interval[b] of -55 to -6 percent of C2 mean.	had no detectable effect on report of non-countable earnings during 2012–2015. Confidence interval[b] of -50 to 6 percent of C2 mean.
Participation Analysis			
Proportion who used the benefit offset by December 2016	3.7 percent	15.8 percent	15.4 percent
Proportion who used BOND work incentives counseling by December 2016	5.0 percent	38.8 percent	95.9 percent
Proportion who used state vocational rehabilitation (VR) services by December 2015	3.8 percent	12.2 percent	15.2 percent [EWIC increased use of VR services relative to WIC and relative to current law]
Proportion with TTW assigned by December 2015	7.0	19.5	25.8 [EWIC increased use of TTW relative to WIC and relative to current law.]
Proportion reporting unmet need for employment services in 36-month survey	37 percent	50 percent	48 percent
Knowledge of benefit offset rules about 3 years after random assignment	29 percent correctly understood that SSA would reduce, but not suspend their benefits because of substantial earnings,7 percentage points larger than for control subjects who wrongly believed that the same was true. 54 percent of Stage 1 control subjects correctly understood current law rules.	48 percent correctly understood that SSA would reduce, but not suspend their benefits because of substantial earnings, 11 percentage points larger than for control subjects who wrongly believed that the same was true. 54 percent of Stage 2 control subjects correctly understood current law rules.	52 percent correctly understood that SSA would reduce, but not suspend their benefits because of substantial earnings, 15 percentage points larger than for control subjects who wrongly believed that the same was true. [EWIC increased knowledge of offset rules relative to WIC by 4 percentage points.]
Proportion with overpayment, mean size of overpayment	3.4 percent with overpayment in any month during 2011–2015. Offset increased the proportion with overpayments by 0.7 percentage points, and decreased mean total overpayment by 28 percent of C1 mean (C1 mean = $348.	13.8 percent with overpayment in any month during 2012–2015.	13.5 percent with overpayment in any month during 2012–2015.
		Offset increased the proportion by 4.7 percentage points, and decreased mean total overpayment by 27 percent of the C2 mean (C2 mean = $1,030).	
Process Analysis			
Work incentives counseling	WIC providers implemented counseling according to design.	WIC providers implemented counseling according to design.	EWIC providers met benchmarks for enhancements to WIC.
Administering the benefit offset	***SSA typically made the first benefit adjustment under the offset many months after the first offset use (i.e., the first month for which benefits were adjusted).***		

Indicator	Target Population and Intervention		
	Stage 1 Full SSDI Caseload	Stage 2 SSDI-Only Volunteers	
	The Benefit Offset (T1)	The Benefit Offset (T21)	The Benefit Offset + EWIC (T22)[a]
	SSA adjusted SSDI benefits a median of 22.2 months after first offset use.	SSA adjusted SSDI benefits a median of 17.3 months after first offset use.	SSA adjusted SSDI benefits a median of 13.2 months after first offset use.
Benefit-Cost Analysis			
Society as a whole[c]	The benefit offset generates net costs to society.	The benefit offset applied to volunteers generates net benefits to society.	The benefit offset applied to volunteers generates net losses to society.

Notes:

[a] Findings do not statistically significantly differ from findings in the "The Benefit Offset (T21)" column, except where noted in brackets.

[b] The confidence interval may be interpreted as follows: if we replicated the demonstration many times under the same general conditions, and each time calculated an impact estimate and a 95 percent confidence interval with the same procedures, then the long-run average of the impact point estimates would be within the confidence interval in 95 percent of the replications.

[c] Findings do not consider the benefits and costs of potential induced entry into SSDI.

8.4.1.   The Benefit Offset Increased SSDI Benefits and Did Not Affect Earnings

For both Stage 1 and Stage 2 of BOND, the evaluation finds confirmatory evidence that the benefit offset led to higher SSDI benefits relative to current law. In contrast, for neither Stage 1 nor Stage 2 do we find confirmatory evidence of an effect of the benefit offset on total earnings.

For the nationally representative cross-section of the SSDI population under age 60 as of May 2011, the Stage 1 impact analysis finds an essentially zero net effect on total earnings over the five-year follow-up period (2011 to 2015). The impact estimate is precisely estimated and the 95 percent confidence interval rules out any policy relevant impact. Underlying this zero net effect, the analysis detects small, theoretically-predicted behavioral responses to the benefit offset policy in opposite directions. The positive impacts on employment and number of years with earnings above BYA indicate increases in earnings for some, while the negative impacts on number of years with earnings two times BYA and three times BYA indicate decreases in earnings for others. The effects nearly cancel each other out resulting in the measured impact on mean total earnings that is essentially zero (less than 0.2 percent).

The magnitude of the impact on SSDI benefits in Stage 1 is $665 over 56 months, or $12 per month (1 percent of the control group mean). Underlying this positive impact on SSDI benefits are factors that changed benefits in opposite directions. On the one hand, two factors increase average SSDI benefits: (i) the windfall to those who would already engage in substantial gainful activity under current law and (ii) the reduction in earnings by some of those beneficiaries receiving windfall income. On the other hand, the increase in the proportion with earnings above BYA decreased average SSDI benefits. As addressed in Chapter 7, by a rough estimate, the Stage 1 effect on the proportion with earnings above BYA would need to be 31 times as large as observed to yield a zero impact on SSDI benefits.

The Stage 2 results show that beneficiaries who volunteered for the demonstration experienced impacts on employment and earnings above BYA that were larger than the typical Stage 1 subject. The statistically insignificant point estimates of impact on total earnings for the two treatment groups represent 9 to 10 percent of the control group average. The 95 percent confidence intervals are 0 percent to 18 percent for the offset plus WIC and -5 percent to +25 percent for the offset plus EWIC. Because the

level of earnings in the control group is low, the upper bounds of these intervals imply an earnings increase of $113 to $154 per month for Stage 2 treatment subjects with any employment during the follow-up period.[150]

The statistically insignificant effects of the offset on earnings in Stage 2 were not enough to produce a reduction in total SSDI benefits. Rather, the average amount of SSDI benefits due to the Stage 2 treatment subjects was about 4 percent larger than the amount due to Stage 2 control subjects. By a rough estimate, the impact on the proportions with earnings above BYA would need to be 8 to 10 times as large as those observed for Stage 2 to yield a zero impact on SSDI benefits.

The impact analysis finds no evidence that enhancements to counseling services under EWIC had an incremental effect on the offset's impacts on earnings or benefit outcomes. Thus, although enhanced counseling had small, positive impacts on understanding of the offset and on use of VR and EN services under the Ticket to Work program, and the participation analysis finds that T22 subjects who used the offset experienced shorter adjustment times and fewer overpayments compared to T21 subjects, these improved process outcomes do not translate into differences in earnings or benefits. The process analysis found that the demonstration implemented the enhanced counseling as intended and that EWIC met the benchmarks for service delivery established in the BOND design.

8.4.2.   The Benefit Offset Has a Net Social Cost for the Full SSDI Caseload

The benefit-cost analysis found a net social cost of the BOND offset for the full SSDI caseload examined in Stage 1. The very small estimated increases in earnings were not sufficient to offset the deadweight loss from increases in taxes to fund larger transfer payments. Distributional effects were much larger, with SSDI beneficiaries gaining income by receiving larger SSDI benefits and countervailing losses to the Disability Insurance Trust Fund. The benefit-cost analysis for Stage 2 shows that the offset policy combined with standard work incentives counseling has a net benefit to society in the small subpopulation of beneficiaries who volunteer for the demonstration. In contrast, the offset policy combined with enhanced work counseling has a net social loss, a result largely due to higher counseling costs.[151]

8.4.3.   Offset Use was Low in Stage 1 and Stage 2

The beneficiaries who use the offset are a combination of those who would have earned more than BYA under current law and those who are induced by the incentive to earn more than BYA. As of December 2016, SSA had adjusted the benefits of 3.7 percent of Stage 1 treatment subjects according to the benefit

---

[150]   The upper bounds of the confidence intervals are $62 and $85 per month for the T21 and T22 impacts, respectively. The $113 and $154 amounts are generated by dividing these monthly amounts by the 55 percent of Stage 2 treatment subjects with any employment during the follow-up period.

[151]   The net benefit to society from the offset plus WIC compared to current law results from a 9 percent earnings increase that more than pays for the deadweight loss of higher taxes to fund increased transfer payments. The net loss to society from the offset plus EWIC compared to current law is largely driven by high costs of counseling. The use of volunteers in Stage 2 means that this finding is not directly relevant to the benefits and costs of a benefit offset applied to the entire SSDI caseload. (Again, this analysis does not consider the benefits and costs of induced entry, which would increase deadweight loss and hence costs to society.)

offset rules.[152] The corresponding percentages for the two Stage 2 treatment groups are 15.8 percent (T21) and 15.4 percent (T22). The impact analysis shows that *offset users are mostly those who would have earned above BYA in the absence of the benefit offset policy*. The evaluation finds that EWIC did not increase offset use compared to WIC services. For Stage 1 subjects, younger age and SSDI-only status are predictive of offset use. Younger age is also predictive of offset use among the Stage 2 sample of volunteers, as are several baseline predictors from survey data including educational attainment and self-reported good to excellent health.

The number of Stage 1 and Stage 2 offset users grew each year of the demonstration. Some of this growth may simply be due to use by those who, at the outset of the demonstration, still had unused TWP and GP months. Some subjects may also have needed time to prepare for and obtain work that would lead to offset use. In addition, the economic expansion over the period may have provided more opportunities to use the offset as the demonstration progressed. Collectively, relative to offset use in the early years of the demonstration, offset use in the later years is more indicative of what use would be under a national program, assuming similar economic conditions. In 2015, 2.2 percent of Stage 1 treatment subjects and 9.7 percent of Stage 2 treatment subjects used the offset.

8.4.4.   There are Several Possible Explanations for Limited Offset Use and Earnings Impacts

The findings from the BOND evaluation raise several questions. Why was offset use not higher among BOND treatment subjects? Why did the benefit offset not evoke a larger behavioral response as evidenced by impacts on earnings and on the proportion of beneficiaries who earned more than BYA? The evidence from the process and participation analyses suggests multiple possible explanations for limited offset use and impacts on earnings but does not answer these questions conclusively. One possible explanation is the limited *work capacity* of most beneficiaries. A second possible explanation is that the increase in the *incentive* to earn more than BYA with the $1 for $2 benefit offset was not strong enough to induce some beneficiaries who can earn more than BYA to actually do so. A third possible explanation stems from the *complexity* of both the intervention and the current law rules. The complexity of the rules may have muddled treatment subject understanding of the change in incentive available through the offset. A final possible explanation rests with the BOND *implementation.* Delays in benefit adjustment and limitations of outreach to Stage 1 beneficiaries suggest that the implementation may have led to smaller impacts than would be the case under a permanent, national benefit offset policy. The next sections discuss evidence from the evaluation supporting these competing explanations.

Beneficiaries' work capacity is limited
All SSDI beneficiaries were determined by SSA to have sufficiently severe impairments that they were unable to engage in SGA for at least 12 months. Hence, it is not surprising that a large share is unable to achieve and sustain earnings above BYA.

The offset's incentive is not strong enough to induce more work effort for some who could do so
Working more requires giving up time spent on other activities, and for some the implicit cost of giving up those activities—the opportunity cost—may be too high, despite the more favorable earnings rules.

---

[152]   These figures are reflected in data available for this report. As in previous years, we expect that SSA will identify additional treatment subjects who used the offset by December 2016 as additional work CDRs are completed after the data extracted for this report.

The $1 for $2 offset for earnings above BYA is the equivalent of a 50 percent marginal tax rate on earnings between BYA and the level of earnings at which benefits fall to zero. Hence, a beneficiary who values the activities that would be displaced at 50 percent or more of the potential increase in gross earnings from working more would presumably not choose to earn more. Taxes on earnings and implicit taxes from reductions in other benefits (for example, SNAP and private disability benefits) further reduce the incentive to earn more than BYA under the offset.

Knowledge of offset rules and current law was limited
Both current law and the benefit offset rules are complex, which results in confusion among beneficiaries about how earnings affect SSDI benefits. Despite the availability of counseling to all beneficiaries, understanding of how earnings affect benefits under current law is low (54 percent for the nationally representative Stage 1 control group provided survey responses consistent with a correct understanding of current law).

Understanding of the offset rules among those subject to them was also low. Outreach in Stage 1 was deliberately limited. Specifically, the implementation team sent two letters to Stage 1 treatment subjects, attempted two phone contacts, and SSA sent one notice. These efforts to inform Stage 1 treatment subjects led to 29 percent of the Stage 1 treatment group demonstrating correct knowledge of how earnings affect benefits under the offset rules three years after random assignment. The relatively limited understanding among the Stage 1 treatment group of how earnings affect benefits under the offset likely constrains the behavioral response to the incentive to at least a small degree.

Stage 2 subjects received outreach and recruitment materials about the benefit offset and completed an informed consent and enrollment process. The evidence shows that compared to Stage 1 the more extensive outreach in Stage 2 led to wider, though still limited understanding of offset rules among Stage 2 treatment subjects. About half of the Stage 2 treatment subjects (48 percent of T21 and 52 percent of T22) provided survey responses consistent with a correct understanding of the offset. Compared to the standard level of counseling, enhanced counseling resulted in a statistically significant, but not large increase in the percentage with a basic understanding of the offset rules (52 percentage points compared to 48 percentages points, i.e. an 8 percent increase).

Benefit Adjustment Involved Long Delays in BOND
Processing of benefit adjustments was problematic. In both stages, the process analysis found that two factors—a backlog in SSA's work CDR processing and functionality problems in the software supporting benefit adjustments—led to substantial delays in the application of offset rules for beneficiaries whose earnings exceeded BYA. Median duration from first month of offset use to first benefit adjustment was 22 months for Stage 1 and 15 months for Stage 2. For Stage 2, enhanced counseling led to shorter times to first benefit adjustment compared to standard work incentives counseling, most likely because of the proactive outreach of EWIC staff. Delayed first adjustments may reduce beneficiary understanding of the offset, and increases the incidence of overpayments.

Delays in adjusting benefits led to overpayments. Although strictly comparable statistics are not available for control subjects, other findings suggest that delays for treatment subjects were typically lengthier than for delays under current law. Likely as a result, the BOND offset increased the prevalence of overpayments in both stages. Despite higher prevalence, because the amount of overpayment in the typical month with an overpayment is much less under the offset rules than under current law; the BOND offset reduced the average size of overpayments.

Offset use was accompanied by overpayments for about 90 percent of offset users. Comparison of overpayments across the two Stage 2 treatment groups found no evidence that EWIC services affected the prevalence of overpayments relative to WIC services. However, we did find evidence that EWIC services reduced the mean total overpayment amount by 23 percent of the offset-plus-WIC mean.

8.5.   Conclusion

The impact analysis finds evidence that the benefit offset increased employment and the share with earnings above BYA in both Stage 1 and Stage 2. Nevertheless, these earnings increases were not large enough in either stage to drive a reduction in average SSDI benefits for all beneficiaries.

It is possible that the impact on the proportion earning above BYA might have been somewhat larger in the nationally-representative Stage 1 had outreach to treatment subjects been more robust and benefit adjustments been made in a more timely manner. However, there is nothing in the evidence to suggest that the magnitude of the effect under an established national program would reach the effect found for volunteers in Stage 2 (which at 23 percent is more than three times greater than the 7 percent relative effect found in Stage 1). Further, the results from Stage 2 imply that even a 23 percent effect would fall far short of what is required to reduce total SSDI benefits—even before consideration of induced entry. Therefore, the evidence from BOND conclusively shows that a national policy that reduces SSDI benefits by $1 for every $2 in earnings above the substantial gainful activity threshold would not reduce the total amount of SSDI benefits owed to beneficiaries.

References

Autor, David H., and Mark G. Duggan. 2006. "The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding." *Journal of Economic Perspectives* 20 (3):71-96.

Bakija, Jon. 2016. "Documentation for a Comprehensive Historical U.S Federal and State Income Tax Calculator Program: Version 2016.0." Williams College, Williamstown, MA 01267.

Bell, Stephen H., and Larry L. Orr. 1994. "Is Subsidized Employment Cost Effective for Welfare Recipients? Experimental Evidence from Seven State Demonstrations," *Journal of Human Resource*19 (1): 42–61.

Bell, Stephen H., Daniel Gubits, David Stapleton, David Wittenburg, Michelle Derr, Arkadipta Ghosh, and Sara Ansell. 2011. "BOND Implementation and Evaluation. Evaluation Analysis Plan." Final Report Submitted to Social Security Administration. Cambridge, MA: Abt Associates..

Bloom, Howard B., and Charles Michalopoulos. 2013. "When Is the Story in the Subgroups? Strategies for Interpreting and Reporting Intervention Effects for Subgroups." *Prevention Science* 14 (2):179-188.

Bloom, Howard S., Carolyn J. Hill, and James A. Riccio. 2003. "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments." *Journal of Policy Analysis and Management*, 22 (4): 551-575.

Boardman, Anthony E., David H. Greenberg, Aldan R. Vining, and David L. Weimer. 2018. *Cost-Benefit Analysis: Concepts and Practice*. 4th ed. New York: Cambridge University Press.

Bureau of Labor Statistics. September 18, 2018. "Employer Costs for Employee Compensation – June 2018." Available at https://www.bls.gov/news.release/ecec.nr0.htm.

Burke, David. June 2016. "Social Security Disability Insurance: Participation and Spending." Congressional Budget Office. Accessed September 1, 2017. Available at https://www.cbo.gov/sites/default/files/114th-congress-2015-2016/reports/51443-ssdiparticipationspending-onecol-2.pdf.

Burkhauser, Richard V., Andrew J. Houtenville, and David C. Wittenburg. 2003. "A User's Guide to Current Statistics on the Employment of People with Disabilities." In *The Decline in Employment of People with Disabilities: A Policy Puzzle,* edited by David C. Stapleton and Richard V. Burkhauser. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. . See esp. pp. 23-86.

Burkhauser, Richard V., and Mary C. Daly. 1996."Employment and economic well-being following the onset of a disability: the role for public policy." In *Disability, Work and Cash Benefits*, edited by Jerry L. Mashaw, Virginia Reno, Richard V. Burkhauser, and Monroe Berkowitz. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. See esp. pp. 59–101.

Card, David, Jochen Kluve, and Andrea Weber. 2015. "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations." Working Paper No. 21431, Cambridge, MA: National Bureau of Economic Research..

Chambless, Catherine E., George Julnes, Sara T. McCormick, and Anne Reither. 2011. "Supporting work effort of SSDI beneficiaries: Implementation of benefit offset pilot demonstration." *Journal of Disability Policy Studies* 22(3): 179-188.

Congressional Budget Office. June 2016. *Social Security Disability Insurance: Participation and Spending.* Available at https://www.cbo.gov/sites/default/files/114th-congress-2015-2016/reports/51443-ssdiparticipationspending-onecol-2.pdf.

Congressional Budget Office. April 2017. *Comparing the Compensation of Federal and Private-Sector Employees, 2011 to 2015.* Washington, D.C.: U.S. Congressional Budget Office. Available at https://www.cbo.gov/system/files?file=115th-congress-2017-2018/reports/52637-federalprivatepay.pdf.

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, David Wittenburg. 2017. "BOND Implementation and Evaluation: 2017 Stage 1 Interim Process, Participation, and Impact Report." Submitted to the Social Security Administration (contract deliverable 24c2.2 under Contract SS00-10-60011), Office of Research, Demonstrations, and Employment Support. Cambridge, MA: Abt Associates.

Dean, David H., Robert C. Dolan, and Robert M. Schmidt. 1999. "Evaluating the vocational rehabilitation program using longitudinal data: Evidence for a quasiexperimental research design." *Evaluation Review* 23 (2): 162-189.

Dean, David, John Pepper, Robert Schmidt, and Steven Stern. 2015. "The Effects of Vocational Rehabilitation for People with Mental Illness." *Journal of Human Resources* 52 (3): 826-858.

Derr, Michelle, Denise Hoffman, Jillian Berk, Ann Person, David Stapleton, Sarah Croake, Christopher Jones, and Jonathan McCay. February 2015. "BOND Implementation and Evaluation: Process Study Report." Submitted to the Social Security Administration (contract deliverable 24d under Contract SS00-10-60011), Office of Research, Demonstrations, and Employment Support. Cambridge, MA: Abt Associates.

Edwards, Ashley. 2015. "Measuring Work-Related Expenses in the Redesigned 2014 SIPP Panel: Methods and Implications." Working Paper No. 723 presented at the February 2016 Eastern Economic Association Meeting. Available at https://census.gov/content/dam/Census/library/working-papers/2015/demo/SIPP-WP-273.pdf.

Evans, William N., and Timothy J. Moore. 2011. "The short-term mortality consequences of income receipt." *Journal of Public Economics,* vol. 95(11): 1410-1424.

Fichtner, Jason J. and Jason S. Seligman. 2016. "Beyond All or Nothing: Reforming Social Security Disability Insurance to Encourage Work and Wealth." In *SSDI Solutions: Ideas to Strengthen the*

Social Security Disability Insurance Program, edited by Jim McCrery, and Earl Pomeroy. West Conshohocken, PA: Infinity Publishing. See esp. Chapter 13.

Fujiwara, Daniel. 2010. "The Department for Work and Pensions Social Cost-Benefit Analysis Framework: Methodologies for Estimating and Incorporating the Wider Social and Economic Impacts of Work in Cost-Benefit Analysis of Employment Programmes." Working paper No. 86 for the Department for Work and Pensions.

Geyer, Judy, Daniel Gubits, Stephen Bell, David Judkins, and Utsav Kattel. 2018a. "BOND Implementation and Evaluation: Fourth-Year Snapshot of Earnings and Benefits Impacts for Stage 2."Submitted to the Social Security Administration (contract deliverable 24c.6 under Contract SS00-10-60011), Office of Program Development & Research. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research.

Geyer, Judy, Daniel Gubits, Stephen Bell, Tyler Morrill, Denise Hoffman, Sarah Croake, Katie Morrison, David Judkins, and David Stapleton. 2018b. "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.

Gibson, Sarah, Greg Mills, Michelle Wood, and David Long. 2008. "Social Security Administration $1 for $2 Benefit Offset Demonstration." Submitted to the Social Security Administration (contract deliverable 21.5 under Contract SS00-04-60110). Cambridge, MA: Abt Associates.

Gokhale, Jagadeesh. 2013. "A New Approach to SSDI Reform; More People Faced with Disabilities Would Return to Work If They Faced Better Incentives." *Regulation* 36(Fall): 44-51.

Greenberg, David H. 1997. "The Leisure Bias in Cost-Benefit Analyses of Employment and Training Programs." *Journal of Human Resources* 32 (2): 413-439.

Greenberg, David H. and Philip K. Robins. 2008. "Incorporating Nonmarket Time into Benefit-Cost Analyses of Social Programs: An Application to the Self-Sufficiency Project." *Journal of Public Economics* 92 (3): 766-794.

Greenberg, David H., Charles Michalopoulos, and Philip K. Robins. 2003. "A Meta-Analysis of Government-Sponsored Training Programs." *Industrial and Labor Relations Review*57 (1): 31-53.

Gubits, Daniel, Judy Geyer, Denise Hoffman, Sarah Croake, Utsav Kattel, David Judkins, Stephen Bell, and David Stapleton. 2017. "BOND Implementation and Evaluation: 2015 Stage 2 Interim Process, Participation, and Impact Report." Submitted to the Social Security Administration (contract deliverable 24c2.3 under Contract SS00-10-60011), Office of Program Development & Research. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research.

Gubits, Daniel, Michelle Derr, Jillian Berk, Ann Person, David Stapleton, Denise Hoffman, Stephen Bell, Rachel Cook, and David Wittenburg. 2013. "BOND Implementation and Evaluation: Stage 2 Early Assessment Report." Submitted to the Social Security Administration (contract deliverable

24a.2 under Contract SS00-10-60011), Office of Program Development & Research.Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research.

Gubits, Daniel, Winston Lin, and Stephen Bell. 2014. "BOND Implementation and Evaluation: First- and Second-Year Snapshot of Earnings and Benefit Impacts for Stage 2." Submitted to the Social Security Administration (contract deliverable 24c.5 under Contract SS00-10-60011), Office of Program Development & Research. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research.

Hanson, Robert Harold. 1978. "The Current Population Survey—Design and Methodology." Technical Paper 40, Washington, D.C.: U.S. Bureau of the Census.

Hamermesh, Daniel S. 1993. *Labor Demand*, Princeton, NJ: Princeton University Press..

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. 2017. "2016 Stage 1 Interim Process, Participation, and Impact Report." Submitted to the Social Security Administration (contract deliverable 24c2.1 under Contract SS00-10-60011), Office of Program Development & Research.. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research.

Isaacs, Julie. 2008. *The Costs of Benefit Delivery in the Food Stamp Program: Lessons from a Cross-Program Analysis.* Washington D.C., Economic Research Service, United Department of Agriculture. Available at www.ers.usda.gov. Accessed November 6, 2017.

Jacobson, Neil, Aya Aghabi, Barbara Butz, and Anita Aaron. 2015. "Exploring an Alternative Social Security Definition of Disability." In *SSDI Solutions: Ideas to Strengthen the Social Security Disability Insurance Program*. See esp. Ch. 14.

Johnson, Nicholas, Phil Oliff, and Erica Williams. 2011. "An Update on State Budget Cuts." Washington, DC: Center on Budget and Policy Priorities.

Laughlin, Lynda. April 2013. "Who's Minding the Kids? Child Care Arrangements: Spring 2011." Working Paper No. 70-135, Washington D.C.: United States Census Bureau.

Layard, R., S. Nickell, and G. Mayraz. 2008. "The Marginal Utility of Income." *The Journal of Public Economics* 92(8) :1846-1857.

Liu, Su, and David C. Stapleton. 2011. "Longitudinal Statistics on Work Activity and Use of Employment Supports for New Social Security Disability Insurance Beneficiaries." *Social Security Bulletin* 71 (3): 35-60.

Livermore, Gina A., Denise Hoffman, and Maura Bardos. 2012. "Ticket to Work Participant Characteristics and Outcomes under the Revised Regulations." Submitted to the Social Security Administration, Office of Retirement and Disability Policy. Washington, DC: Mathematica Policy Research.

Maestas, Nicole, Kathleen J. Mullen, and Gema Zamarro. 2010/ "Research Designs for Estimating Induced Entry into the SSDI Program Resulting from a Benefit Offset." Technical Report TR-908-SSA. Santa Monica, CA: RAND Labor and Population.

Mashaw, J. L., and Reno, V. P, eds. 1996. "Balancing security and opportunity: The challenge of disability income policy". Washington, DC: National Academy of Social Insurance.

Michalopouos, Charles, David Wittenburg, Dina A.R. Israel, Jennifer Schore, Anne Warren, Aparajita Zutshi, Stephen Freedman and Lisa Schwartz. February 2011. "The Accelerated Benefits Demonstration and Evaluation Project: Impacts on Health and Employment at Twelve Months. Volume 1" New York, NY: MDRC.

Morton, William R. 2013. "Social Security Disability Insurance (SSDI) Reform: An Overview of Proposals to Reduce the Growth in SSDI Rolls." Washington, D.C.: Congressional Research Service..

Newhouse, Joseph P. and Rand Corporation. 1993. Insurance Experiment Group. *Free for all?: Lessons from the RAND health insurance experiment*. Cambridge, MA: Harvard University Press.

Rehabilitation Research and Training Center on Disability Statistics and Demographics. 2009. *Annual Compendium of Disability Statistics: 2009*. U.S. Department of Education, National Institute on Disability and Rehabilitation Research, grant no. H133B080012. New York, NY: Hunter College.

Robins, Philip K., and R.W. West. 1980. "Labor Supply Response over Time." *Journal of Human Resources*15(4): 524-544.

Schimmel Hyde, Jody, and David C. Stapleton. 2015. "Changes to the Ticket to Work Regulations in 2008 Attracted Providers and Participants, But Impacts on Work and Benefits Are Unclear." *Social Security Bulletin*75(4): 15-33.

Schimmel, Jody, Sarah Prenovitz, Gina Livermore, and Alex Bryce. 2013. "Evaluation of the Work Incentives Planning and Assistance (WIPA) Program in 2011: Beneficiaries Served, Services Provided, and Program Costs." Submitted to the Social Security Administration. Washington, DC: Mathematica Policy Research.

Schochet, Peter Z. 2009. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Evaluation Review*33 (6): 539-567.

Social Security Administration, Office of the Inspector General. 2015. *Overpayments in the Social Security Administration's Disability Programs—A 10-Year Study.* Baltimore, MD: SSA, 2015. Available at https://oig.ssa.gov/sites/default/files/audit/full/pdf/A-01-14-24114.pdf..

Social Security Administration, Office of the Inspector General. 2017 *Overpayments Collected Through Long-term Repayment Plans*. Baltimore, MD: SSA, 2017. Available at https://oig.ssa.gov/sites/default/files/audit/full/pdf/A-07-16-50082.pdf.

Social Security Administration. January 26, 2011. "The Social Security Administration's Report on Continuing Disability Reviews Resulting from Work Activity." Report to House Appropriations Committee. Washington, DC: SSA..

Social Security Administration. 2013. *OASDI and SSI Program Rates & Limits, 2014.* Washington, DC: SSA.

Social Security Administration. 2016a. *The 2016 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds.* Available at https://www.ssa.gov/oact/tr/2016/tr2016.pdf.

Social Security Administration. 2016b. *Annual Statistical Supplement to the Social Security Bulletin, 2015.* Washington, D.C.: SSA, 2016.

Social Security Administration. 2017a. *Annual Statistical Supplement to the Social Security Bulletin, 2016.* Washington, D.C.: SSA, 2017.

Social Security Administration. 2017b. *The 2017 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds.* Washington, D.C.: SSA, 2017. Available at https://www.ssa.gov/oact/tr/2017/tr2017.pdf

Social Security Administration. 2017c. "Changes in Program Policy Influencing Program Size." Washington, D.C.: SSA, 2017. Available at https://www.ssa.gov/policy/docs/chartbooks/disability_trends/sect04.html.

Social Security Administration. 2017d. "2017 Red Book: SSDI and SSI Employment Supports." Washington, D.C.: SSA, 2017. Available at https://www.ssa.gov/redbook/eng/ssdi-and-ssi-employments-supports.htm. Social Security Administration. 2018. *The 2018 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds.* Available at https://www.ssa.gov/oact/tr/2018/tr2018.pdf.

Social Security Administration. n.d. "Average CPI By Quarter And Year." Available at https://www.ssa.gov/oact/STATS/avgcpi.html. Accessed November 16, 2017.

Social Security Administration. n.d. "Demonstrations: What are demonstration projects?" Available at https://www.ssa.gov/disabilityresearch/demos.htm. Accessed August 30, 2017.

Social Security Administration. n.d. "Promoting Opportunity Demonstration." Available at https://www.ssa.gov/disabilityresearch/pod.htm. 2017. Accessed November 16, 2017.

Social Security Administration. n.d. "VR Reimbursement Claims Processing." Available at https://www.ssa.gov/work/claimsprocessing.html. Accessed June 19, 2018.

Social Security Advisory Board. A Disability System for the 21st Century. Washington, DC: SSAB, 2006.

Stapleton, David C., Stephen H. Bell, David C. Wittenburg, Brian Sokol, and Debi McInnis. December 2010. "BOND Implementation and Evaluation: BOND Final Design Report." Submitted to the

Social Security Administration (contract deliverable 2.2 under Contract SS00-10-60011), Office of Program Development & Research. Cambridge, MA: Abt Associates.

Stapleton, David, David Wittenburg, Daniel Gubits, David Judkins, David Mann, and Andrew McGuirk. 2013. "BOND Implementation and Evaluation: First-Year Snapshot of Earnings and Benefit Impacts for Stage 1." Submitted to the Social Security Administration (contract deliverable 24c.3 under Contract SS00-10-60011), Office of Program Development & Research. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research.

Stapleton, David, David Wittenburg, David R. Mann, Denise Hoffman, and Andrew McGuirk. 2014. "BOND Implementation and Evaluation: Second-Year Snapshot of Earnings and Benefit Impacts for Stage 1." Submitted to the Social Security Administration (contract deliverable 24c.2 under Contract SS00-10-60011), Office of Program Development & Research. Washington, DC: Mathematica Policy Research.

Stapleton, David C., and Jody Schimmel Hyde. 2017. "Employment Support for the Transition to Retirement." *Research on Aging* 39(1) : 249-271.

Tremblay, Tim, James Smith, Alice Porter, and Robert Weathers. 2011. "Effects on Beneficiary Employment and Earnings of a Graduated $1-For-$2 Benefit Offset for Social Security Disability Insurance (SSDI)." *Journal of Rehabilitation* 77(2): 19.

Tuma, Nancy B. 2001. "Approaches to Evaluating Induced Entry Into a New SSDI Program with a $1 Reduction in Benefits for Each $2 in Earnings." Washington, DC: Social Security Administration.

United States Census Bureau. n.d. "Table 2B: Mean Monthly Earnings by Selected Characteristics: 2012." Microsoft Excel spreadsheet. Washington, D.C.: U.S. Census Bureau. Available at www2.census.gov/programs-surveys/sipp/2008-panel/2012-annual/earnings/2012-ann-earn-table2B.xls.

Walczak, Jared and Scott Drenkard. 2017. "State and Local Sales Tax Rates in 2017." Washington D.C.: Tax Foundation. Available at https://taxfoundation.org/state-and-local-sales-tax-rates-in-2017.

Weathers III, Robert R. and Jeffrey Hemmeter. 2011. "The impact of changing financial work incentives on the earnings of Social Security Disability Insurance (SSDI) beneficiaries." *Journal of Policy Analysis and Management* 30 (4): 708-728.

Westfall, Peter H., Randall Tobias, and Russell D. Wolfinger. 2011. *Multiple Comparisons and Multiple Tests Using SAS*. Cary, NC: SAS Institute, 2011. See esp. pp. 349-350.

Westfall, Peter H., S. Stanley Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.* New York: John Wiley & Sons, Inc.

Winkler, R. L. 2003. *An Introduction to Bayesian Inference and Decision.* 2nd ed. Sugarland, TX: Probabilistic Publishing.

Wittenburg, David, David R. Mann, David Stapleton, Daniel Gubits, David Judkins and Andrew McGuirk. 2015. "BOND Implementation and Evaluation: Third-Year Snapshot of Earnings and Benefit Impacts for Stage 1." Submitted to the Social Security Administration (contract deliverable 24.c.3 under Contract SS00-10-60011), Office of Program Development & Research.. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research.

Wittenburg, David, David Stapleton, Michelle Derr, Denise W. Hoffman, and David R. Mann. May 2012. "BOND Implementation and Evaluation: BOND Stage 1 Early Assessment Report." Submitted to the Social Security Administration (contract deliverable 24.1 under Contract SS00-10-60011), Office of Program Development & Research. Cambridge, MA: Abt Associates.

Wolter, Kirk. 1985. *Introduction to Variance Estimation*. New York: Springer-Verlag.

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