Caution to readers: The estimates produced from IRS earnings and SSA benefit paid data in this report were later updated to include refinements to the analytic methodology and data. The specific variables affected are: Total earnings, Employment, Earnings above BYA, Earnings above 2XBYA, Earnings above 3XBYA, Total SSDI benefits paid, Number of months with SSDI payments, Total SSI benefits paid, and Number of months with SSI payments. The data and statistical methods used to produce these estimates have been updated over the course of the demonstration, making the published estimates in this report out of date. For the most up-to-date estimates, please refer to the Final Evaluation Report which will be available in late 2018.

BOND Implementation and Evaluation

2016 Stage 1 Interim Process, Participation, and Impact Report

Deliverable 24c2.1

March 20, 2017

Submitted to:
Social Security Administration
Attn: Ms. Joyanne Cobb
Office of Research, Demonstration and Employment Support (ORDES)
500 E Street, SW
9th Floor, RM #905
Washington, DC 20254

Contract No. SS00-10-60011

Prepared by:
Denise Hoffman
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
Michael Sinclair

Table of Contents

List of Exhibits

Acronyms Used in This Report

AEE	Annual Earnings Estimate
BODS	BOND Operations Data System
BOND	Benefit Offset National Demonstration
BPP	BOND Participation Period
BS&A	Benefits Summary and Analysis
BSAS	BOND Stand Alone System
BTS	Beneficiary Tracking System
BYA	BOND Yearly Amount
CDR	Continuing Disability Reviews
CPE	Centralized Post-Entitlement
CWIC	Community Work Incentive Coordinators
DAC	Disabled Adult Child
DAF	Disability Analysis File
DBAD	Disabled Beneficiary and Dependent
DWB	Disabled Widow/Widowers Benefits
EN	Employment Network
EPE	Extended Period of Eligibility
ESA	Employment Success Advisor
EWIC	Enhanced Work Incentives Counseling or Counselor
EXR	Expedited Reinstatement
FTE	Full-Time Equivalent
GP	Grace Period
I&R	Information and Referral
IRP	Initial Reinstatement Period
IRS	Internal Revenue Service
MEF	Master Earnings File
MBI	Medicaid Buy-In
MBR	Master Beneficiary Record
ORDES	Office of Research, Demonstration and Employment Support
PHUS	Payment History Update System
SGA	Substantial Gainful Activity
SNAP	Supplemental Nutrition Assistance Program
SSA	Social Security Administration
SSDI	Social Security Disability Insurance
SSI	Supplemental Security Income
SSR	Supplemental Security Record
SVRA	State Vocational Rehabilitation Agency
TANF	Temporary Assistance for Needy Families
TWP	Trial Work Period
UI	Unemployment Insurance
WIC	Work Incentives Counseling or Counselor
WIP	Work Incentives Plan
WIPA	Work Incentives, Planning, and Assistance

Executive Summary

As part of the Ticket to Work 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 their reliance on benefits. In response, SSA has undertaken the Benefit Offset National Demonstration (BOND), a random assignment test of variants of SSDI program rules governing work and other supports. BOND incorporates a $1 for $2 benefit offset that allows beneficiaries to retain more of their monthly cash benefit while working.

The BOND project includes two stages. Stage 1 supports an evaluation of how a national $1 for $2 benefit offset would affect earnings and program outcomes for the entire SSDI population. In contrast, Stage 2 supports a more in-depth evaluation of impacts on those beneficiaries more likely to use the offset (recruited and informed volunteers) and an investigation of the degree to which enhanced counseling services—focused on work incentives—affect impacts.

This report, the first of two Stage 1 Interim Process, Participation, and Impact Reports, uses several data sources to document findings from Stage 1. The report tracks process and participation findings using a combination of service-use data and interviews with BOND subjects to document events through 2015. It also includes a review of the administrative processes associated with implementing the offset and related outcomes, such as overpayments following a retroactive benefit adjustment, based on administrative data and interviews with staff involved in these activities. An impact analysis compares outcomes for all Stage 1 treatment subjects (T1 subjects) with those for all Stage 1 control subjects (C1 subjects), using administrative data through 2014 and a 2014 survey of random samples of T1 and C1 subjects, to determine if the BOND intervention affects SSDI benefits, earnings, and other outcomes in 2014, the most recent year with data.

Summary of Key Findings

1. Offset Usage and Beneficiary Knowledge of Offset Rules

Since the opportunity to use the offset began in May 2011, the share of T1 subjects who have ever used benefits counseling services or the offset has grown gradually. Through December 2015, nearly 5 percent of T1 subjects had used Work Incentive Counseling (WIC), the T1 benefits counseling services comparable to those available under current law to C1 subjects. Of those, nearly 80 percent had used services beyond information and referral services. Over the same period, SSA had adjusted the benefits of 2.8 percent of T1 subjects under the offset. Approximately half of these T1 offset users entered the offset following proactive reporting, as SSA requires. The other half entered the offset through the SSA-initiated automated reconciliation process that takes place after the end of the calendar year for all treatment subjects known to have completed the Trial Work Period (TWP) and Grace Period (GP) and subsequently engaged in SGA.

T1 subjects' knowledge and understanding of the offset rules and how benefits relate to earnings was limited, even among those most likely to use the offset. In the survey, 29 percent of T1 subjects provided a correct answer to a question about how earnings affect their benefits under the offset. A marginally higher portion of T1 subjects who were employed the year before the demonstration—a group for whom

the offset rules had greater salience—demonstrated an understanding of the benefit offset's earnings rules. By comparison, 54 percent of C1 subjects provided the correct current-law answer.

2. Administrative Processes and Overpayments

Delays in the benefit-adjustment process are substantial and have persisted throughout the demonstration. For all offset users with first adjustments in 2013 through 2015, the median duration from the first month in which a reduced benefit payment under the offset was due until the first month in which SSA applied the adjustment was 22 months as of December 2015. This represents a significant portion of the 60-month period during which treatment subjects may use the offset. The median duration of the delay was shorter (17 months) for those who entered the offset following proactive reporting of earnings.

For both T1 and C1 subjects, delays in adjustment of benefits following TWP and GP completion often resulted in overpayments. Relative to C1 subjects, T1 subjects are more likely to accrue overpayments, but they have lower overpayment amounts. Among T1 offset users from 2011 through 2013, 83 percent had received an overpayment. The incidence of overpayments over all T1 subjects was 0.23 percentage points higher than for C1 subjects, representing a 12 percent increase relative to the C1 mean of 1.88 percent. Given that the amount of an overpayment in a month with an overpayment is typically a fraction of the full monthly benefit amount for T1 subjects, while C1 subjects are overpaid by their full benefit amounts, the mean overpayment made to T1 subjects was lower than that made to C1 subjects, despite the higher incidence. If we assume that all C1 subjects who were overpaid also would have been overpaid had they been assigned to BOND, then over the first three years of BOND the mean overpayments of T1 subjects with overpayments were about $3,100 less than they would have been under current law. Because the stated assumption might not be exactly correct, this should be considered an upper bound for the point estimate of the mean reduction in the size of overpayments for those who would have an overpayment under the BOND benefit offset, current law, or both.

A noticeably high share of the 20 offset users who participated in in-depth interviews also reported overpayments. A quarter of these same interviewees identified the overpayment as a negative feature of their demonstration experience, including two who reported resulting financial hardships.

3. Earnings and Benefit Outcomes (Administrative Data)

Of the several earnings and benefit outcomes examined, the analysis plan for BOND (Bell et al. 2011) identifies two confirmatory outcomes as the most important in assessing the demonstration's overall impacts: annual earnings and total SSDI benefits paid in a year. Statistical tests for the occurrence of these impacts use a multiple-comparisons procedure to compensate for what would otherwise be an elevated risk of false positive findings. We also explore impacts on other earnings and benefit outcomes in the administrative data, without making multiple-comparison adjustments of this sort—thus increasing the risk of false positive findings for this portion of the analysis.

We find no evidence that the BOND benefit offset had an impact on the total earnings of T1 subjects in 2014. We did find strong evidence that it had a positive impact on SSDI benefits paid in the same year. The magnitude of the impact on SSDI benefits paid in 2014 was $132, or $11 per month.

Findings from additional exploratory analyses indicate more nuanced effects on other earnings outcomes that are consistent with theoretical predictions. The estimated positive effect of BOND on the percentage

employed is 2.2 percent of the control group percentage, and the estimated positive impact on the percentage earning more than the BOND Yearly Amount (BYA; the earnings threshold above which SSA applies the benefit offset) is 7.4 percent of the control group percentage. These findings, both of which are statistically significant, are consistent with the theoretical prediction that the BOND offset will increase employment and the percentage of beneficiaries earning more than BYA. We also find evidence of negative impacts on the percentage of beneficiaries earning above three times BYA, consistent with the prediction that some higher-earning T1 subjects will reduce their earnings in order to receive a partial benefit under the offset.

1. Introduction

As part of the Ticket to Work and Work Incentives Improvement Act of 1999, Congress asked the Social Security Administration (SSA) to test alternative Social Security Disability Insurance (SSDI) work rules that are designed to increase the incentive for SSDI beneficiaries to work and reduce their reliance on SSDI benefits. In response, SSA has undertaken the Benefit Offset National Demonstration (BOND), a random assignment test of variants of SSDI program rules governing work and other supports. SSA, in conjunction with several contractors led by Abt Associates, developed the infrastructure and supports required to implement BOND.

The BOND project includes two stages. Stage 1 is designed to examine how a national benefit offset would affect earnings and program outcomes for the entire SSDI population. Stage 2 is designed to learn about impacts for those more likely to use a benefit offset—volunteers recruited from the SSDI-only population (those SSDI beneficiaries not also receiving Supplemental Security Income, or SSI)—and to determine the impacts of the addition of more intensive counseling around work and benefits to an offset.

This report, the *2016 Stage 1 Interim Process, Participation, and Impact Report*, documents results of the Stage 1 process and participation analyses into the fifth calendar year of implementation (2015). This report also documents impacts on earnings and benefit outcomes during the fourth calendar year of implementation (2014), approximately contemporaneous outcomes measured in a survey of Stage 1 subjects, and impacts on overpayments—when SSA pays beneficiaries more or less than they were entitled and later reconciles the difference—through 2013.

Three *Stage 1 Snapshot Reports* (Stapleton et al. 2013; Stapleton et al. 2014; Wittenburg et al. 2015) have documented Stage 1 impacts on earnings and benefit outcomes during the first three calendar years (2011, 2012, and 2013) of implementation. Future reports—a second *Stage 1 Interim Process, Participation, and Impact Report* and the *Final Report*, both scheduled for 2017—will track Stage 1 impacts through 2016. We are producing a parallel series of reports for Stage 2, the first of which was released in 2014.

This introductory chapter describes the specific benefit offset implemented by BOND and Stage 1 of the demonstration (Section 1.1). The chapter also reviews the objectives of the BOND evaluation and the research questions addressed by the process, participation, and impact analyses (Section 1.2). Section 1.2 also summarizes primary findings to date on the implementation and impacts of BOND as documented in previous reports. The chapter concludes by describing the organization of the remainder of the report (Section 1.3).

1.1. The BOND Policy Test

Under current program rules, SSDI beneficiaries lose all SSDI benefits after a sustained period of substantial earnings and risk the potential loss of other (non-SSDI) benefits.[1] Specifically, beneficiaries

[1] Other benefits include Medicare for those on the rolls for at least 24 months. These benefits are extended for a long 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 may be reduced or eliminated as earnings increase.

lose SSDI benefits if their countable monthly earnings exceed the monthly Substantial Gainful Activity (SGA) amount after completing a nine-month Trial Work Period (TWP) and a three-month Grace Period (GP). In 2014, the year for which Stage 1 impacts are analyzed in this report, the SGA amount was $1,070 per month for non-blind beneficiaries and $1,800 per month for blind beneficiaries. The complete loss of benefits for sustained earnings in excess of the SGA amount is sometimes called the "cash cliff." Economic theory predicts that the cash cliff discourages some beneficiaries from working at all and encourages those who work and could earn above the SGA level to keep their earnings below that level.

BOND replaces the cash cliff with a "ramp" (benefit offset), with the policy objective of encouraging beneficiaries who can work above the SGA level to increase their earnings and reduce their reliance on benefits.[2] More specifically, benefits decrease by $1 for every additional $2 in countable earnings above an annualized version of SGA once the beneficiary has exhausted the SSDI program's TWP and GP. By protecting partial benefits for those who earn at this level, the benefit offset is expected to increase earnings for some beneficiaries who otherwise might not work at all or would earn less than the SGA amount. If the offset induces such individuals to earn above SGA, their benefits will be reduced. However, higher benefits will be paid to some beneficiaries who would have earned above the SGA amount and received no SSDI cash benefit. Further, for two reasons, such beneficiaries may choose to reduce their earnings somewhat; first, they will have higher incomes, and, second, any reduction in the amount of earnings above SGA will produce an increase in SSDI benefits equivalent to half the earnings reduction. Thus, the direction of the average impact on mean earnings and benefits of all beneficiaries will depend on the size of the impacts for beneficiaries who would not engage in SGA under current law, relative to the size of the impacts for those who would engage in SGA under current law.

BOND also changes the administrative processes used to adjust benefits and 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 2014, $12,840 for non-blind and $21,600 for blind beneficiaries).[3] The benefit offset reduces benefits by $1 for every $2 in countable annual earnings in excess of the BYA following completion of the GP. The change to an annual period can also help beneficiaries who have variable monthly earnings. SSA continues to pay benefits monthly under BOND, but the monthly payment amount is based on expected annual earnings. In the following calendar year, SSA reconciles payments to actual countable earnings, based on information provided by the Internal Revenue Service (IRS), documentation provided by the beneficiary, or both.

Beneficiaries eligible for the benefit offset may use it during the 60-month BOND Participation Period (BPP), which begins the month after demonstration entry for those who completed the TWP before that month or in the month after a given beneficiary's TWP ends, provided that the TWP is completed by September 30, 2017. Those who do not complete the TWP by that date will lose their opportunity to use the offset. SSA will not permanently terminate benefits because of work during the BPP, even if benefits fall to zero because of earnings. SSA will apply current rules at the end of the BPP and will terminate the benefits of those engaged in SGA in any month after the last GP month.

---

2 All references in this report to the benefit offset refer to the offset as implemented in BOND.

3 Stage 1 impacts documented in this report are analyzed through calendar year 2014. In 2015, the BYA was $13,080 for non-blind beneficiaries and $21,840 for blind beneficiaries.

As noted, BOND includes two stages. Stage 1 tests the impact of the benefit offset on the overall SSDI population.[4] Stage 2 examines the offset's impact on a group that consists of individuals who are more likely to use the offset than the average SSDI beneficiary, that is, volunteers recruited from the SSDI-only population and informed about the offset before enrollment and random assignment. Exhibit 1-1 illustrates the process by which the demonstration formed groups to test the offset. The current report concerns Stage 1, which has the goal of examining how a national benefit offset and changes to ancillary supports would affect earnings and program outcomes for the entire SSDI population. To that end, the demonstration randomly selected 10 large study sites to statistically represent the nation. These sites are Alabama, Arizona/Southeastern California, Colorado/Wyoming, DC Metro, Greater Detroit, Greater Houston, Northern New England, South Florida, Western New York, and Wisconsin. A computer routine randomly assigned beneficiaries in the sites to either a treatment group that receives the benefit offset (T1 subjects) or a control group that continues under standard rules (C1 subjects). By design, T1 and C1 subjects were to have access to counseling regarding work and benefits of a roughly similar level of intensity. C1 subjects were to have access to counseling under an existing program—Work Incentives Planning and Assistance (WIPA). T1 subjects were to have access to similar counseling services, customized to the special rules that apply to their benefits, called Work Incentives Counseling (WIC).

By virtue of random assignment, the T1 and C1 groups as a whole should be statistically equivalent, so that any statistically significant differences in outcomes between T1 and C1 subjects can be confidently attributed to the intervention—following the basic impact measurement strategy in a randomized experiment. The final evaluation sample—beneficiaries randomly assigned to Stage 1 of BOND—includes 79,436 T1 subjects and 901,709 C1 subjects.[5]

Also using an experimental design, Stage 2 aims to learn about the impacts of the benefit offset on those beneficiaries more likely to use it—informed volunteers recruited from the SSDI-only population—and to determine the effects of the delivery of more intensive counseling services, called Enhanced Work Incentives Counseling (EWIC), relative to current law and relative to WIC services. To achieve these goals, Stage 2 uses three-way random assignment into an offset-plus-WIC group (T21 subjects), an offset-plus-EWIC group (T22 subjects), and a current-law benefits group (C2 subjects). In total, the Stage 2 sample includes 12,744 beneficiaries. Concurrent beneficiaries—SSDI beneficiaries who also were receiving SSI at the time of random assignment—were excluded from Stage 2 because the interaction between the SSI work incentives and the benefit offset (which applies to SSDI benefits) substantially diminishes the value of the benefit offset to such beneficiaries.[6]

---

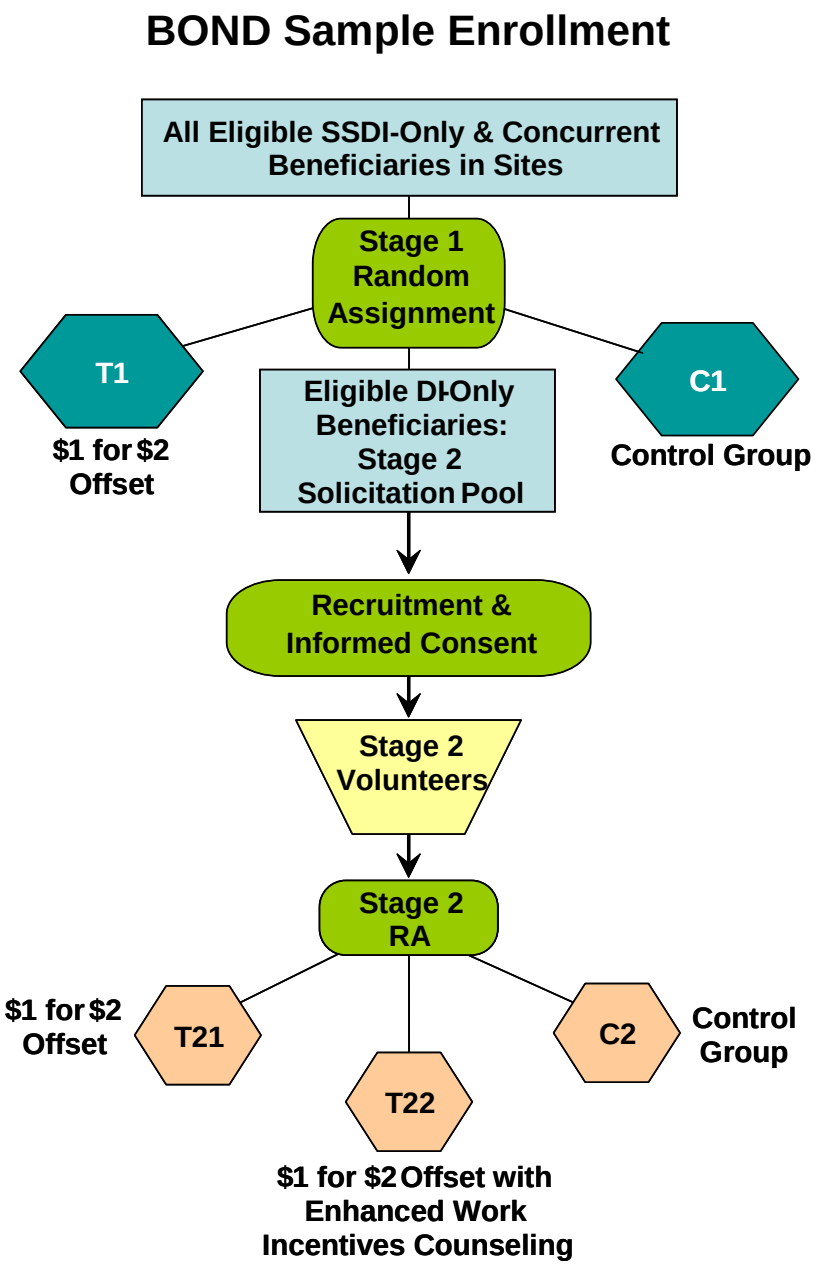
4 Throughout this report references to the impact of the benefit offset, are inclusive of the benefit offset, counseling, and administrative components of the BOND intervention available to treatment subjects.

5 The final Stage 1 evaluation sample excludes beneficiaries who were randomly assigned but later identified as having died before random assignment. Compared to the evaluation sample, the final Stage 1 analysis sample used for the impact analysis further excludes pairs of related beneficiaries who receive disability benefits based on a common primary beneficiary's record if the two members of the pair were assigned to different Stage 1 groups (T1 versus C1). See Section 2.2.3 for more details.

6 See the *Evaluation Analysis Plan* (Bell et al. 2011) for more details on the random assignment process and the reasons for excluding concurrent beneficiaries from Stage 2 but not from Stage 1.

This report examines interim results of the evaluation of Stage 1 of the demonstration. For the Stage 1 evaluation, we refer to the combination of the $1 for $2 offset, the annual measure of SGA, the associated administrative changes to the benefit adjustment process, and the availability of WIC services for benefits counseling as the benefit offset.

Exhibit 1-1. Overview of BOND Random Assignment Process



1.2. The BOND Evaluation

Abt Associates, in partnership with Mathematica Policy Research, is conducting a comprehensive evaluation of the BOND interventions, including studies of demonstration implementation, beneficiary participation, net impacts, and net social costs and benefits. The evaluation will include cross-cutting analyses that combine findings from these four components. These studies draw on the various components of the evaluation to deepen our understanding of how the BOND interventions affected beneficiaries. Initial findings from the process, participation, and impact analyses for the demonstration's two stages appear in earlier reports as shown in Exhibit 1-2.

This section describes the research questions addressed by the process, participation, and impact analyses and summarizes key findings documented in earlier reports.

The BOND Evaluation Team

Abt Associates, in partnership with 25 other organizations, is implementing and evaluating 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. Neither these individuals nor any member of their team has a role in implementing or overseeing the BOND intervention they are studying. Separation of implementation and evaluation does not extend throughout the project, however. The Abt Project Director (Michelle Wood) and Principal Investigator (Howard Rolston) have joint responsibility 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.

Exhibit 1-2. Earlier 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) • *Process Study Report* (Derr et al. 2015)	• *Stage 2 Early Assessment Report* (Gubits et al. 2013) • *Process Study Report* (Derr et al. 2015) • *2015 Stage 2 Interim Process, Participation, and Impact Report* (Gubits et al. 2016)
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)	• *First- and Second-Year Snapshot of Earnings and Benefit Impacts for Stage 2* (Gubits et al. 2014) • *2015 Stage 2 Interim Process, Participation, and Impact Report* (Gubits et al. 2016)

1.2.1. The Process Analysis

The overarching objective of the process analysis is to document the characteristics of the BOND intervention, creating a foundation for interpreting impacts. To that end, the process study evaluates the implementation of BOND within and across the study sites over time and assesses the fidelity of the implementation relative to the original design. The process study includes seven rounds of field work activities over the course of the demonstration and relies on several data sources, including feedback from beneficiaries. As described in the *Evaluation Analysis Plan* (Bell et al. 2011), the process analysis uses a mix of qualitative and quantitative data to address five broad research questions:[7]

1. How was the intervention implemented for Stage 1 and Stage 2? How did the implementation evolve over time?
2. Were the recruitment and enrollment processes for Stages 1 and 2 implemented as designed? If significant deviations occurred, why did they occur?
3. Were WIC and EWIC services implemented as designed? To what extent did EWIC services differ from WIC services?
4. Were the processes for reporting earnings, determining TWP completion, and making benefit adjustments for Stages 1 and 2 implemented as designed? How well did the processes perform?
5. What are the likely implications for demonstration outcomes? What are the lessons for national implementation of a benefit offset?

Summary of Findings to Date on the Implementation of BOND

Two reports detail the evaluation's early process analysis findings for Stage 1. The *Stage 1 Early Assessment Report* (Wittenburg et al. 2012) covered the period through November 2011, seven months after random assignment and enrollment of Stage 1 participants in April 2011.[8] The *Process Study Report* (Derr et al. 2015) reported on implementation through the third calendar year of the demonstration (2013), including changes that occurred since November 2011. This section describes key findings from both reports.

BOND Infrastructure and Operations. As reported in the *Stage 1 Early Assessment Report*, the quick start-up of this complex and multifaceted demonstration was a considerable challenge for the BOND Implementation Team. Notably, sample selection and random assignment in spring 2011 produced treatment (T1) and control (C1) groups that were well matched at baseline. The main tasks during the initial months of the demonstration included building the BOND infrastructure (for example, hiring and training WIC/EWIC staff and obtaining security clearances); defining policies and procedures; designing and testing the BOND Operations Data System (BODS); and organizing and executing outreach efforts. The *Process Study Report* discussed changes to BOND policies and procedures that had occurred since those earliest months, including, most significantly, a shift in responsibilities for preparing information

[7] The Evaluation Team has made slight modifications to these research questions compared to the version stated in the *Evaluation Analysis Plan*.

[8] Stage 1 random assignment occurred in late April 2011 and beneficiaries who had already completed the TWP and GP were eligible to use the BOND offset starting in May 2011.

for work Continuing Disability Reviews (work CDRs, which are used to track completion of the TWP and subsequent earnings). The initial demonstration design called for BOND field staff to conduct work CDR preparation activities and submit needed information to SSA. In May 2012, the responsibility shifted to SSA staff.

In addition, some aspects of the BOND infrastructure were not yet functioning as well as intended at the end of November 2011. The BOND Evaluation Team documented challenges associated with coordinating tasks, the competing demands on limited resources, and rapidly changing policies and procedures. In particular, challenges during the initial implementation period included delays in obtaining security clearances for implementation staff, time needed to acquire laptops and other technology, changes to the Beneficiary Tracking System (BTS, a component of BODS), and policy and procedural changes such as when to initiate a work CDR to determine if benefits need to be adjusted. However, the Evaluation Team later concluded that BOND implementation had gradually improved since inception so that, as of fall 2013, the demonstration was largely functioning as designed.

Counseling Available to Control Group. The *Process Study Report* documented an important change in how counseling services to BOND control group subjects and all non-BOND beneficiaries were provided partway through the demonstration. Nationally, WIPA funding expired on June 30, 2012, with no indication that it would be reinstated. More than a year later, in August 2013, Congress resumed WIPA funding. Changes to the WIPA program created some disruptions in counseling services for C1 subjects, though SSA continued to provide some WIPA-like services through the Ticket to Work Help Line and Community Work Incentive Coordinators (CWICs). In BOND sites where demonstration agencies also served as WIPA providers, these changes to the WIPA program also created disruptions in BOND staffing.

Outreach and Enrollment. Evidence described in the *Stage 1 Early Assessment Report* indicated that, even though initial Stage 1 outreach efforts were executed as designed, some beneficiaries were confused about the demonstration. For example, field staff reported that beneficiaries stated that the outreach letters they received were unclear or that they initially disregarded the letters. Field staff also discussed how some local SSA field offices and disability service providers that were not aware of BOND cautioned beneficiaries about contacting demonstration staff because they thought BOND might be a scam. The Evaluation Team heard reports of confusion about BOND from multiple sources and sites. We were unable, however, to quantify the extent to which beneficiaries received misinformation about BOND or how well beneficiaries understood the information they received through the initial outreach effort.

In an attempt to increase awareness about the benefit offset and counseling services, the Implementation Team subsequently made additional outreach attempts to Stage 1 beneficiaries by letter and telephone, as described in the *Process Study Report*. In 2012, the team targeted 10,388 T1 subjects for outreach, most of whom had earnings in 2011. Then, in 2013 and early 2014, the team targeted the remaining 60,345 T1 subjects who had not already been in contact with the demonstration. Immediately following these efforts, the numbers of subjects contacting the demonstration, contacting their assigned WIC counselor, and proactively starting the benefit adjustment process increased by noticeable amounts, indicating that additional outreach increased T1 subjects' awareness of the demonstration's services and requirements. However, most T1 subjects did not respond to any outreach efforts or were not reached because of outdated or inaccurate contact information. The Evaluation Team could not determine the extent to which

the limited response to outreach efforts reflected lack of interest in the benefit offset rather than the outreach efforts' failure to adequately inform those who might find the offset of substantial interest.

Pathway to the Offset. As reported in the *Stage 1 Early Assessment Report*, through October 2011 SSA had adjusted the benefits of 21 T1 subjects, signifying that the subjects had entered the benefit offset. The 21 T1 subjects accounted for a small fraction of the 4,840 T1 subjects who had been in contact with the demonstration at the time. However, based on benefit adjustments made by SSA through May 2014, 0.9 percent of T1 subjects (695 beneficiaries) had used the offset by the end of 2011. The difference suggests that most early offset users did not proactively engage with the demonstration to allow for timely adjustment. Instead of following this "front-door" path to the offset, many subjects later entered through a "back-door" path after SSA became aware of their substantial earnings from sources external to the demonstration, primarily reports of earnings from the IRS.

In the *Process Study Report*, the Evaluation Team found that the number of subjects using the benefit offset was growing steadily over time. Based on benefit adjustments made through May 2014, SSA had identified 1.5 percent of T1 subjects (1,159 beneficiaries) as offset users by the end of 2013. The Evaluation Team concluded that use of the offset was likely to continue rising as SSA retroactively adjusted the benefits of offset users and more treatment subjects engaged in SGA and qualified for the offset.

The Evaluation Team also found that, since the initial implementation of BOND, lengthy delays with benefit adjustments under the benefit offset had been common and, in many cases, led to overpayments for beneficiaries. Given that such lags are common under current law, BOND control subjects presumably also experienced delays. Most causes of lags in benefit adjustment are the same under the offset and current law. Under both sets of rules, delays can occur because (1) beneficiaries do not report earnings in a timely manner (to SSA field offices or, in the case of the offset, to the demonstration); (2) backlogs occur in tasks such as processing work CDRs; or (3) beneficiaries are slow to respond to requests for information when SSA starts the adjustment process. In addition, the following three considerations suggest that the demonstration itself likely contributed to the lags among treatment subjects: (1) the need to inform treatment subjects about the offset; (2) start-up problems in the post-entitlement processes (such as calculating Annual Earnings Estimates) that are needed to facilitate benefit adjustments; and (3) delayed completion of adjustments following SSA review of IRS data on BOND treatment subjects' earnings.

The process analysis documented improvements in the timeliness of benefit adjustment under BOND over time, particularly a decrease in work CDR processing times. Substantial delays remained at the end of 2013, but the Evaluation Team expected that the increase in available resources to process work CDRs and the centralization of post-entitlement work would reduce delays even further.

This report presents updated process analysis findings through 2015, the fifth of six calendar years of implementation covered by the BOND evaluation reports. Findings include new information from recent in-depth, one-on-one interviews with work-oriented T1 subjects.

1.2.2. The Participation Analysis

The participation analysis documents the engagement of prospective BOND subjects in work activities throughout the demonstration.[9] The Evaluation Team expected that T1 subjects' use of BOND demonstration services would vary across beneficiary subgroups (e.g., younger versus older beneficiaries). The Stage 1 participation analysis summarizes patterns of participation by subgroups and more broadly examines whether the BOND interventions influenced all types of work activity, including the use of demonstration services. The participation analysis examines the following questions:

1. To what extent do treatment subjects work or use employment services and benefits counseling?
2. Who works, uses counseling services and other work incentives, and eventually uses the benefit offset?
3. How does the demonstration affect the use of work incentive counseling and the services delivered by counselors?
4. What characteristics distinguish offset users from others?
5. How do work and use of work incentives vary across demonstration groups?
6. How do work and use of work incentives change with time?

Summary of Findings to Date from the Participation Analysis

Both the *Stage 1 Early Assessment Report* and *Process Study Report* contain early results from the participation analysis, specifically Stage 1 participation in WIC. In the *Stage 1 Early Assessment Report*, the Evaluation Team found that 1,024 Stage 1 beneficiaries used WIC services as of October 2011, representing just over 1 percent of all T1 subjects and 21 percent of the 4,840 T1 subjects who had been in contact with the demonstration to that point. WIC service use was presumably well below the capacity of the demonstration, which was designed to accommodate contacts by up to 30 percent of T1 subjects by September 2017. Of the 1,024 T1 WIC users, 334 (33 percent) received basic information and referral services. The remaining 690 users (67 percent) received more intensive WIC services, such as obtaining in-depth information on benefit receipt and work history so that counselors could provide specific counseling on work incentive use, earning goals, and needed employment and personal supports.

In the *Process Study Report*, the Evaluation Team reported that the number of beneficiaries using WIC services increased to 4,413 as of January 2014. This number includes both T1 and T21 subjects, the latter of whom are part of Stage 2 of the demonstration. However, given that the number of T21 WIC users was smaller than the number of T1 WIC users, the number of T1 subjects using WIC services clearly increased by a large amount since 2011. In analyzing the additional T1 outreach—the extra letters sent and telephone calls made to increase awareness of the benefit offset and counseling services between 2012 and early 2014—the report noted that 2.5 percent of the first group of 10,388 targeted T1 subjects and 1.1 percent of the remaining group of 60,345 T1 subjects contacted their WIC for services during the approximate period of the outreach to their respective group. Altogether, almost 1,000 more T1 subjects started using WIC services during the additional T1 outreach effort.

[9] For Stage 2, there is an additional component of the participation analysis, which focuses on recruitment of volunteers into the demonstration and responses by various subgroups of beneficiaries.

This report presents updated findings about Stage 1 beneficiary engagement in benefits counseling and, for the first time, beneficiary engagement in work and related activities and receipt of employment supports. The report also examines benefit offset use over time and by beneficiary characteristics.

1.2.3. The Impact Analysis

The Stage 1 impact analysis addresses the following question:

1. What would be the impact of the benefit offset on outcomes for all SSDI beneficiaries nationally as compared to their outcomes under current benefit payment rules?

Impacts may come from a variety of sources. First, impacts may be related to the $1 for $2 offset, the annual accounting period, and the associated administrative changes. Specifically, the administrative procedures established to provide T1 subjects with information and to implement benefit adjustments under the offset likely affected the speed with which SSA made payment adjustments. Given how they are measured, these adjustments are especially important for the estimated impacts on benefits paid. Due to the time frame of this report, we could not include data that became available after the end of 2015. Therefore, this report estimates impacts on benefits paid *in* 2014. Because of retroactive benefit adjustments and resulting over- and underpayments, impacts on benefits paid *for* 2014 might differ substantially. The *Final Report* will include estimates of the impact of BOND on benefits paid *for* the years in the evaluation period, as opposed to impacts for benefits paid *in* those years.

Finally, T1 subjects have access to counseling services that are tailored to the benefit offset but are otherwise intended to be comparable to counseling services available to all beneficiaries under current law and hence offered to C1 subjects. It is possible, though not intended, that the implementation of the counseling services offered to T1 subjects differs from that offered to C1 subjects in ways that have an impact on earnings and benefits above and beyond the impact of the offset itself.

Summary of Findings to Date from the Impact Analysis

The *Third-Year Snapshot of Earnings and Benefit Impacts for Stage 1* (Wittenburg et al. 2015) documented Stage 1 impacts on earnings and benefit outcomes during the third calendar year of implementation (2013). The impact estimates showed that the benefit offset, as administered under BOND, did not have a statistically significant impact on total earnings in 2013—similar to previously reported findings for 2011 and 2012.[10] In addition, the offset did not have a statistically significant increase on the proportion of T1 subjects earning above BYA.

[10] Findings for 2011 and 2012 first appear in the *First-Year Snapshot of Earnings and Benefit Impacts for Stage 1* (Stapleton et al. 2013) and *Second-Year Snapshot of Earnings and Benefit Impacts for Stage 1* (Stapleton et al. 2014), respectively. For 2013, the *Third-Year Snapshot of Earnings and Benefit Impacts for Stage 1* used a new method to compute standard errors in order to improve their stability. In that report, we also recalculated the 2011 and 2012 results by using the new method and found that any differences obtained from applying the new method were minimal and did not substantively change the interpretation of findings from the earlier reports. See the *Third-Year Snapshot* for more information.

We also reported a positive impact on total SSDI benefits paid *in* 2013; the point estimate for the 2013 impact is twice the size of the 2012 impact estimate. As discussed earlier in this section, the impact on benefits paid *in* a given year—such as 2013—does not represent the impact on benefits that will be eventually paid *for* that year; in this case, benefits paid *in* 2013 do not reflect retroactive adjustments to benefit payments made after 2013, including adjustments for the benefit offset. The previously noted absence of an impact on the percentage with earnings above BYA suggests that the impact on benefits paid *for* 2013 will be positive. That is because the offset can reduce benefits paid for a given year only by inducing beneficiaries to earn above BYA when they would not otherwise have done so. For beneficiaries who would earn above BYA under current law, benefits increase because beneficiaries become eligible for partial benefits under the offset.

In the current report, we present the impact findings for the same outcomes in 2014, the fourth calendar year of implementation. For the first time, we also report impacts on additional outcomes in several domains, including attitudes toward employment, work-related expenses, health status, time use, and overpayments.

1.3. Organization of the Current Report

The remainder of this report is organized into 10 chapters. Chapter 2 describes the data sources and analytic methods used in the report. Chapter 3 updates contextual information from the 2013 process analysis concerning the BOND study sites and disability service environment.

Using updated data from BTS and new data from recent in-depth, semi-structured interviews with 30 work-oriented T1 beneficiaries, Chapter 4 describes benefits counseling in BOND and updates the process and participation findings from previous reports. In particular, we explore the content of benefits counseling and the extent to which T1 subjects received such counseling.

Chapter 5 presents findings from the participation analysis, including analyses of new data from a survey of both T1 and C1 subjects conducted 36 months after demonstration enrollment. The survey collected information on Stage 1 subjects' awareness of BOND and understanding of how earnings affect SSDI benefits. We also present findings from the in-depth interviews of 30 work-oriented treatment beneficiaries.

Chapter 6 analyzes new results involving employment, primarily taken from a survey of Stage 1 beneficiaries and incorporating information from the in-depth interviews of work-oriented treatment subjects.

Chapter 7 describes the structure and implementation of the $1 for $2 benefit offset. We also update the process and participation analyses of Stage 1 subjects' use of the offset and benefit adjustment since the 2013 process analysis, including an examination of new statistics on the duration of offset use.

Chapter 8 expands on discussions in earlier reports regarding overpayments. The analysis incorporates information from the in-depth interviews of work-oriented beneficiaries and includes the first estimates of the impact of the benefit offset on the prevalence and size of overpayments.

Chapters 9 and 10 present the large share of the findings from the Stage 1 impact analysis. Chapter 9 focuses on impacts on 2014 earnings and SSDI benefits measured from SSA administrative data,

including variations in impacts between various subgroups of beneficiaries as defined by their background characteristics. Chapter 10 presents impact estimates on additional employment-related outcomes and outcomes in other domains, using new data from a survey of Stage 1 beneficiaries.

Chapter 11 provides a summary of key findings and conclusions to date.

Four appendices provide additional technical details and findings. Appendix A describes the impact estimation methodology for administrative outcomes. Appendix B describes the impact methodology for outcomes derived from the Stage 1 36-Month Survey. Appendix C discusses how we measure overpayments, which are analyzed in Chapter 8. Appendix D presents impact estimates for 2014 for all of the beneficiary subgroup analyses described in Chapter 9. Appendix E provides detailed impact estimates based on survey data, as described in Chapter 10.

2. Methodology

To support findings from the Stage 1 process, participation, and impact analyses, this chapter describes the data sources used in each analysis (Section 2.1). The chapter then presents an overview of the methodology for each analysis, including the approach to estimating impacts of the benefit offset (Section 2.2). Additional methodological detail appears in appendices as cited.

2.1. Data

The analyses in this report are based on data from several sources, including semi-structured interviews and focus groups with BOND staff and T1 subjects, a survey of T1 and C1 subjects, program implementation information from demonstration operations, and SSA administrative data. This section describes the data sources that support each part of the evaluation: the process analysis first, followed by the participation analysis, and finally the impact analysis.

2.1.1. Data Sources for the Process Analysis

The process analysis involves eight rounds of data collection from several sources over the course of the demonstration. The current report uses information from the first seven rounds of data collection covering BOND implementation through 2015. As part of these efforts, the Evaluation Team 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 also 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. To assess BOND implementation, the Evaluation Team also used administrative data from BODS on the delivery of demonstration services and beneficiary status. This section provides an overview of the qualitative data collection activities from the two most recent rounds of data collection, which provide the primary data for the process analysis in this report.[11]

In fall 2014, the study team conducted the sixth round of qualitative data collection, which consisted of two main activities: (1) telephone/online focus groups with WIC and EWIC supervisors and counselors during October 2014 and (2) telephone interviews with key members of the BOND Implementation Team and the SSA BOND operations team in January 2015. Data collection focused on documenting the changes made to BOND since the previous round of data collection in fall 2013 and on staff perceptions of those changes. Data collection also supported the identification of successes, challenges, and lessons in implementing BOND and how they might influence the impact of the demonstration.

The Evaluation Team conducted the seventh round of data collection for the process analysis in fall 2015. Data collection involved two primary activities. First, we conducted informal interviews with members of the Implementation Team and the SSA BOND operations team to learn more about implementation-related changes that took place after the January 2015 interviews. Second, to learn more about factors

[11] Earlier reports describe other earlier data collection efforts. In particular, see the *Process Study Report* for the fifth round of qualitative data collection and the *Stage 1 Early Assessment Report* and *Stage 2 Early Assessment Report* for earlier rounds.

influencing beneficiaries' employment and benefit offset use, we conducted in-depth, semi-structured telephone interviews with work-oriented beneficiaries in September and October 2015.

In the rest of this subsection, we describe the activities conducted for the sixth and seventh rounds of data collection.

Telephone focus groups with WIC and EWIC staff. In fall 2014, the Evaluation Team conducted 10 telephone/online focus groups with 51 WIC and EWIC supervisors and counselors (Exhibit 2-1).[12] Each focus group involved staff from multiple BOND sites. The team organized the groups to collect data separately (1) from sites in which post-entitlement responsibilities had shifted from WIC and EWIC staff to a centralized team versus those in which these responsibilities remained with WIC and EWIC staff;[13] (2) from supervisors versus counselors; and (3) from those involved in WIC versus EWIC services.[14] To identify potential participants, the Evaluation Team asked the Implementation Team to recommend WIC and EWIC supervisors and counselors who could represent their respective roles at each site. To recruit participants, the team sent an email invitation two to four weeks in advance of the focus groups, followed by a telephone call to non-responders, and an email reminder one to two days ahead of the meeting.[15]

During each 90-minute focus group, trained facilitators led the telephone discussion using protocols and conducted an online poll to capture answers to multiple-choice questions. The focus group topics relevant to Stage 1 included staff education and experience; community context and service environment; BOND organizational and staffing infrastructure; roles and responsibilities of WIC staff; perception of centralization (centralized sites only), perception of post-entitlement responsibilities (non-centralized sites only); over and underpayments; influence of BOND on beneficiary behavior; and successes and challenges. During the focus group discussions, the facilitators invited (but did not require) participants to respond to an online, multiple-choice poll.

[12] The 18 WIC non-supervisory counselors who participated in focus groups represented 39 percent of the total 46 counselors at all sites, and the 14 EWIC non-supervisory counselors who participated in focus groups represented 49 percent of the total 29 counselors at all sites.

[13] In this context, post-entitlement services refer to services that facilitate the BOND benefit adjustment process.

[14] In this report, we do not use data relevant only to Stage 2, but we incorporate information that applies to both Stage 1 and Stage 2 beneficiaries (for example, experiences with non-BOND employment services).

[15] Of the supervisors and counselors we contacted, 79 percent (56 of 71) either participated in the focus group or spoke with us individually if they were unable to participate on the scheduled date at the given time.

Exhibit 2-1. Qualitative Data Collection – Fall 2014 WIC/EWIC Focus Groups

	Focus Group	Sites Represented	Number of Focus Group Participants	Live Poll Respondents
Supervisors				
1	WIC Centralized Group (supervisors)	Arizona/SE California, Colorado/Wyoming, DC Metro, Northern New England, South Florida, Wisconsin	8	7
2	WIC Non-Centralized (supervisors)	Greater Detroit, Western New York	2	2
3	EWIC Centralized (supervisors, pilot)	Arizona/SE California, Colorado/Wyoming, Greater Detroit, Greater Houston, South Florida	5	5
4	EWIC Non-Centralized (supervisors)	Alabama, DC Metro, Wisconsin, Western New York	4	3
Counselors				
5	WIC Centralized Group 1 (counselors, pilot)	Arizona/SE California, Greater Houston, Northern New England	3	3
6	WIC Centralized Group 2 (counselors)	DC Metro, Northern New England, South Florida, Wisconsin	10	10
7	WIC Non-Centralized (counselors)	Alabama, Greater Detroit, Western New York	5	5
8	EWIC Centralized Group 1 (counselors)	Arizona/SE California, Colorado/Wyoming	3	3
9	EWIC Centralized Group 2 (counselors)	Northern New England, South Florida	5	5
10	EWIC Non-Centralized (counselors)	Alabama, DC Metro, Wisconsin, Western New York	6	6
Total	**10 groups***	**All BOND sites**	**51**	**49**

*In addition, we conducted five calls with individuals who were unable to participate in the scheduled focus groups, where the absence left a gap in representation for a particular site and role. Specifically, these calls were with Colorado WIC counselor, Northern New England WIC supervisor, Western New York WIC counselor, Greater Houston WIC supervisor, and Greater Houston EWIC counselor. As of November 2014, BOND had a total of 46 WIC counselors and 29 EWIC counselors.

Telephone interviews with SSA and BOND Implementation Team. In January 2015, the Evaluation Team conducted five telephone interviews with nine key members of the BOND operations team at SSA and the Abt-led BOND Implementation Team (which, as noted in Chapter 1, is separate from the Evaluation Team). We interviewed four members of the BOND operations staff from the SSA's Office of Research, Demonstration and Employment Support (ORDES). They are responsible for a variety of tasks, such as overseeing the BOND Stand Alone System (BSAS, a computer program that interfaces with SSA's data systems to adjust SSDI benefits for treatment subjects according to BOND rules) and processing work CDRs. Within the Implementation Team, interviewees included the director and deputy director of implementation, the liaison to all BOND sites, and the lead and another member of the team

providing technical assistance to WIC and EWIC staff and conducting centralized post-entitlement work. We selected team members most familiar with BOND processes, changes to processes, and the reasons for those changes and completed interviews with all identified individuals.

During each 60-minute telephone interview, interviewers used a protocol tailored to the role of each respondent. The interviewers focused on clarifying the information discussed during the staff focus groups and identifying key changes to implementation. Across all five interviews, the topics of discussion relevant to Stage 1 included the BOND service environment; WIC services; WIC organizational performance; work CDRs; centralization of post-entitlement work; Annual Earnings Estimates (AEEs, used by BSAS to prospectively adjust benefits) and mailers collecting information on non-countable earnings; and benefit adjustments, overpayments, and underpayments.

In October 2015, the Evaluation Team held an informal interview with a member of the Implementation Team to update information on BOND service providers, with a focus on changes related to the new round of WIPA grants awarded in 2015 and updated staffing configurations. In December 2015, we interviewed the SSA BOND operations staff at ORDES, who addressed staffing in the ORDES BOND unit; work CDR processing; automated reconciliations; and other changes to BOND policies or processes.

Telephone interviews with work-oriented treatment subjects. Members of the Evaluation Team conducted in-depth, one-on-one telephone interviews with a total of 60 work-oriented BOND treatment subjects. Half of the respondents (30) were T1 subjects and are included as a data source for this report. The remaining respondents were Stage 2 subjects (T21 and T22 subjects); we will incorporate information from these subjects in future reports covering Stage 2.

The primary purpose of the interviews was to gain the perspectives of beneficiaries who appear to be work-oriented with respect to the following questions:

1. Why do some beneficiaries work but not to the point at which they take advantage of the benefit offset?
2. Why do other beneficiaries take advantage of the offset but only for a short time?
3. Why does a third group of beneficiaries use the offset for a long period of time?

For operational purposes, we defined work-oriented BOND subjects as those with a disability cessation date; in other words, SSA had determined that the subjects engaged in SGA after completing their TWP.[16] Subjects who work but not to the point of disability cessation could also be considered work-oriented, but we have no way to use administrative data to distinguish between those within this group for whom the

[16] SSA determines disability cessation dates by conducting work CDRs. These disability cessation dates for work above the SGA level may have occurred several years before the interviews, in some cases even before the beneficiary was randomly assigned to the T1 group in 2011. Despite the elapsed time, we consider such subjects to be work-oriented because their past engagement in work needed to reach disability cessation is evidence of work since becoming a SSDI beneficiary and signals potential interest in using the offset during the demonstration period.

benefit offset would and would not be salient.[17] Among those with disability cessation dates, we identified three subgroups: (1) those that did not use the offset; (2) those who used the offset in a single calendar year between 2011 and 2015 before discontinuing use (short-term offset users); and (3) those who used the offset for three or more consecutive years between 2011 and 2015 (long-term offset users).[18] To identify the three subgroups, we used data extracted from BTS in June 2015, which contains records for all T1 subjects and includes information on cessation dates and whether or not the beneficiary used the offset during each calendar year.

Within each of the three subgroups (those that did not use the offset, short-term offset users, and long-term offset users), we approximated a random sample by calling T1 subjects in the sequence in which they appeared on a randomly ordered list until we completed interviews with 10 subjects. Because we did not make repeated attempts to contact non-respondents to the initial attempt, interviewees in each subgroup may not be representative of all members of the respective subgroup. Nevertheless, we have no reason to think that their responses systematically misrepresent the experiences of their subgroup. We mailed beneficiaries who participated in an interview a $25 check for their time. The interviews lasted about 20 to 30 minutes. Interviewers used a semi-structured guide with several sections: initial reaction to BOND; understanding of BOND and the offset; employment patterns (including key questions about factors affecting work and earnings); experience with the BOND counselor; the benefit adjustment process (if applicable); experience with over- and underpayments; and overall experience with the demonstration.

2.1.2. Data Sources for the Participation Analysis

The participation analysis relies on demonstration operations data, information from the beneficiary survey, and SSA administrative data.

BODS is a data management system designed specifically for BOND. BTS, which is a core component of BODS, includes data documenting contact with the demonstration and participation in the BOND counseling intervention. BODS also includes information obtained from SSA administrative data on whether SSA has determined that disability has ceased because of work above SGA. Such information is used to identify BOND subjects who may be eligible for the benefit offset. In addition, BODS tracks steps associated with benefit offset adjustment.

The BOND Stage 1 36-Month Survey, conducted approximately three years after Stage 1 random assignment, was a telephone survey with field follow-up between May 2014 and February 2015. The survey took the form of a questionnaire requiring on average 40 minutes to administer. Survey content includes information on service receipt, employment and earnings at the time of the survey, receipt of education and training, health and functional status, awareness of BOND or SSA work incentives, and demographic information, among other topics.

[17] The offset is likely to be of most salience to beneficiaries with cessation dates. Although some beneficiaries without a cessation date may be motivated by the offset to work or earn more, we know that a large majority of beneficiaries without a cessation date are not working or have worked relatively little since entering SSDI. Use of readily available cessation dates is a practical way to find subjects for whom the offset is likely to be salient.

[18] To heighten the contrast between short-term and long-term offset users, we did not include those who used the offset for two (and only two) years in any subgroup.

The study team's goal was to complete interviews with at least 80 percent of a sample of 10,000 BOND subjects, evenly split across T1 and C1 subjects and selected to statistically represent the entirety of both groups of subjects. Finding subjects and engaging them in the survey was much more difficult than expected. Ultimately, we achieved a response rate of 59 percent—5,735 interviews. T1 survey subjects completed the survey at a slightly higher rate (59.1 percent) than C1 survey subjects (58.2 percent). The lower-than-expected response rate reduced our power to detect impacts on survey-measured outcomes and increased the risk of non-response bias. To reduce the potential impact of non-response bias on survey measures, both as a whole and separately for T1 and C1 respondents, the survey weights include a non-response adjustment that accounts for observed differences in the rate of survey response across several beneficiary characteristics from administrative data. We describe the non-response adjustment process in detail in Appendix B; the characteristics used in the process appear in Exhibit B-4.[19]

The **Disabled Beneficiary and Dependent (DBAD)** files provide monthly snapshots of SSDI program activity. The files reflect program activity at the time the data were pulled (once per month) rather than the most up-to-date SSA data (which may include retroactive adjustments). SSA updates the DBAD only when there is a change associated with the beneficiary's record. Each snapshot lists up to 35 effective dates and associated actions with each date. The actions apply during the time range of effective date n to effective date n + 1. That is, the information is relevant from the effective date listed until a new effective date appears in a future monthly extract. We exploit documentation of changes in SSA actions over time both across and within DBADs to construct monthly measures of work-related overpayments.[20]

2.1.3. Data Sources for the Impact Analysis

For the impact analysis documented in this report, we used administrative data for benefits and earnings paid *in* calendar year 2014. Benefit outcomes are measured from SSA's Payment History Update System (PHUS) for SSDI and from the Supplemental Security Record (SSR) for SSI. We measured earnings from the SSA Master Earnings File (MEF), which contains longitudinal information on wages and self-employment income reported to the IRS.[21] SSA extracted the MEF records for calendar year 2014 in November 2015. At that time, the calendar year 2014 earnings records were considered nearly 100 percent complete. In addition, we used data from the Stage 1 36-Month Survey to estimate impacts for a multitude of outcomes. Finally, we use data from the DBAD files to estimate impacts on overpayments.

We initially specified the administrative outcomes for the impact analysis in Bell et al. (2011) and every evaluation report has used these specifications. This report uses the same measures, but clarifies that, in

[19] We provide an overview of the sample weights in Section 2.2.3 and detail the process for creating these weights in Appendix B.

[20] Overpayments and incorrect payments occur when SSA pays beneficiaries more SSDI benefits than those to which they are entitled (Section 8.1). Those that are work-related occur as a result of earnings exceeding designated program thresholds.

[21] Because the data are collected by the IRS and are therefore subject to IRS access rules, cleared 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.

all reports, the administrative earnings measure includes only "Social Security earnings." Social Security earnings are earnings that are taxable for Social Security purposes.[22] About 6 percent of the U.S. work force hold jobs not covered by Social Security taxes. Furthermore, Social Security earnings are capped at a maximum taxable amount, $117,000 for 2014. 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 that amount. In 2014, 0.03 percent of Stage 1 subjects had earnings equal to the 2014 maximum taxable amount and 0.01 percent had earnings above the 2014 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 estimate of the impact 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. Measures of weekly earnings and employment taken from survey data are not subject to the same source of bias.

In addition, we used administrative data to develop covariates and subgroups used in the impact analysis. We took the baseline variables used as covariates from the Master Beneficiary Record (MBR) and SSR and took the variables used to form subgroups for the impact analysis from the MBR, SSR, publicly available data on state availability of the Medicaid Buy-In, and MEF.

2.2. Methods

In this section, we describe the methods supporting each type of analysis: process, participation, and impact.

2.2.1. Methods for the Process Analysis

To identify key themes from the 2014 WIC/EWIC staff focus groups and 2015 telephone interviews with work-oriented T1 subjects, the Evaluation Team coded and analyzed responses within and across respondent subgroups. We analyzed beneficiary interview responses as a whole and by the three categories of benefit offset use (those who did not use the offset, short-term offset users, and long-term offset users). For the WIC/EWIC focus groups, we analyzed subgroup responses based on staff role (WIC or EWIC) and site type (centralized or non-centralized post-entitlement work). We also analyzed the online, multiple-choice poll responses across all of the WIC/EWIC focus groups and identified themes within each focus group.

For the in-depth beneficiary interviews and WIC/EWIC focus groups, if we asked all respondents a question, we use counts and percentages to describe their responses. We mention any exceptions where they occur. For example, in some cases, we indicate that a number of respondents commented on a specific topic but that we did not discuss the topic with all respondents.

[22] Appendix A.3 describes the earnings data sources in more detail.

The Evaluation Team reviewed responses from the interviews with SSA and the Implementation Team for details, illustrations, and other information on how BOND was implemented. We used the responses to understand and contextualize findings from the analysis. For example, we used information from interviews with SSA's BOND operations team to understand the factors that caused delays in the adjustment of benefits under the benefit offset after a treatment subject first earns enough to use the offset.

When interpreting any findings from interviews with work-oriented T1 subjects, two central limitations should be considered. First, the respondents and their experiences are not representative of all BOND or all work-oriented T1 subjects. Responses among the three cessation date subgroups might be representative of all T1 subjects within the respective subgroups, but, as mentioned earlier, interviewees may differ in some respects from those who did not respond to our contact attempt. Second, the small sample sizes (10 in each subgroup) mean that sampling error may play a substantial role in determining the reported percentages. Nevertheless, the findings, though not necessarily representative, are useful for explanations and illustrations of subjects' experiences, perceptions, and actions that cannot be obtained from other sources.

2.2.2. Methods for the Participation Analysis

In the participation analysis, the Evaluation Team analyzed several data sources. We used BODS data to create descriptive statistics on WIC counselor caseloads, beneficiary receipt of BOND counseling, the percentage of T1 subjects who completed steps toward benefit adjustment, and the duration of offset use. We used a combination of BODS data and SSA administrative records to identify beneficiary characteristics associated with offset use. We also used MBR data to track the percentage of T1 subjects in the offset over time. Finally, we used the DBAD files to create statistics on overpayments, as described below.

Overpayments

Work-related overpayments and incorrect payments occur when beneficiary earnings exceed thresholds that require SSA to reduce or withhold SSDI benefits but, for a variety of reasons, SSA paid the beneficiary more than he or she was owed. In the remainder of this chapter, we use "overpayments" to refer to work-related overpayments and incorrect payments.[23] Both types of overpayments are included in the statistics presented and overpayments for reasons unrelated to work are excluded.

SSA does not produce readily available statistics indicating the number and amount of overpayments that accrued over a specific period. Rather, SSA's Recovery of Overpayments, Accounting, and Reporting

[23] Conceptually, overpayments and incorrect payments are identical and treated as such by the overpayment identification method. They both apply to cases in which a beneficiary was paid more than he or she was owed. They are distinguished administratively by the payment recovery procedures—an incorrect payment occurs within the accounting period (e.g., the current calendar year for BOND treatment subjects) and is recovered immediately and fully. An overpayment is discovered after the accounting period (e.g., after the end of the calendar year in which the payment error occurred) and is eligible for appeal and repayment arrangements.

system lists overpayments according to when SSA identified the overpayment.[24] Because overpayments are identified with lags of varying lengths, we are unable to use this data system to identify overpayments that accrued during BOND. To address this challenge, we used the DBAD files to develop a method to estimate overpayments that accrued to both treatment and control subjects while in BOND.

The sample included in the overpayment analysis includes only the disabled-worker beneficiaries in the Stage 1 impact samples. The analysis excludes the relatively small number of disabled adult children and disabled-worker beneficiaries assigned to BOND because of difficulties in distinguishing between benefit changes due to their primary beneficiary's earnings from those due to their own earnings. All statistics pertain to overpayments for the disabled-worker's own benefits and do not include overpayments for auxiliary benefits.

We identified overpayments during the first 32 months of BOND for T1 subjects: May 2011 through the December 2013. Even though we are interested in overpayments beyond 2013, we limited our analysis to the specified period because of the often lengthy lag between overpayment occurrence and SSA's discovery of the overpayment. Indeed, because SSA may continue to identify new overpayments as it receives and processes necessary information, the statistics we present are lower-bound estimates of the prevalence of overpayments.

The basic computation of an overpayment is the difference between benefits due as initially recorded for a given beneficiary in that month and benefits due after SSA has received and processed information about work that may retroactively affect benefits. We based the analysis on benefits due (conditional on having been paid a positive benefit) rather than strictly on benefits paid; we believe that the former is a more accurate reflection of work-related benefit adjustments.[25] Specifically, benefits paid exhibit more variance and may reflect variation not related to work. For example, SSA sometimes pays beneficiaries multiple times the amount of benefits due because of decisions applied retroactively for reasons entirely unrelated to earnings.

Appendix C presents additional details about construction of the overpayment measure.

2.2.3. Methods for the Impact Analysis

The central issue in the BOND evaluation is the benefit offset's impact on beneficiary employment, earnings, and benefit receipt. With its elimination of the SGA cash cliff in favor of a benefit offset, SSA intended the offset to create a stronger incentive to increase earnings, on average, and (eventually) lead to reduced benefits and a smaller number of beneficiaries. In this section, we describe how we analyze impacts, including the outcomes of interest, the expected effects of the offset, our impact estimation methodology, and our analysis sample. Appendices A and B present details on the estimation methodology.

[24] Recovery of Overpayments, Accounting, and Reporting does indicate the overpayment accrual dates in a minority of cases. This does not allow us to identify the universe of overpayments accrued during the BOND period.

[25] Benefits due indicate the amount of benefits owed in that month based on activity in that month. Benefits paid are benefits due plus any adjustments, such as withholdings, to repay prior overpayments or SSA lump-sum transfers to reconcile previous underpayments.

Administrative outcome definitions and theoretical impacts

Consistent with past Stage 1 snapshot reports, the current report presents impacts on nine outcomes measured from administrative data, including two confirmatory outcomes (2014 annual earnings and total SSDI benefits paid *in* 2014) and seven exploratory outcomes related to employment and benefits. It is important to note that the confirmatory outcome for benefits in the final report will be different: benefits paid *for* the years in the evaluation period, instead of benefits paid *in* each year. Benefits paid *for* a year will reflect all of the retroactive adjustments that SSA has made since the end of the year through the date on which we extract the data for the final report from SSA records.

The exploratory earning outcomes include indicators for 2014 earnings in excess of each of three annual earning thresholds defined by multiples of BYA (one, two, and three times BYA) and an indicator for any employment during 2014 (defined as positive earnings in 2014). The exploratory benefit outcomes include the number of months with SSDI payments, total SSI benefits paid, and the number of months with SSI payments—each in 2014. In addition, the report includes impact estimates for the prevalence and size of SSDI benefit overpayments among the subset of T1 subjects who are disabled-worker beneficiaries, as defined above.

In Exhibit 2-2, we list the nine administrative outcomes analyzed in all BOND impact reports, provide a definition of each outcome, and indicate the predicted direction of impact, if any (positive, negative, or ambiguous), based on the conventional theory of labor economics as developed in the *Evaluation Analysis Plan* and summarized here. The empirical analysis in later chapters will test for evidence for or against the theory and estimate the magnitude of the impact (and associated standard errors).

Below, we consider the direction of impact on the outcome measures expected from the benefit offset implemented by BOND. Our discussion initially ignores any impact of administrative factors that could influence the outcomes (Section 1.2.3). We then turn to a discussion of administrative factors and their potential influence on impacts.

The goal of BOND is to test how eliminating the SGA cash cliff and replacing it with the $1 for $2 offset ramp (a $1 reduction in benefits for every additional $2 earned) affects return to work, earnings, and beneficiaries' reliance on SSDI benefits. The theoretical direction of impacts of the benefit offset on total earnings[26] and benefits is ambiguous. As described in detail in the *Evaluation Analysis Plan*, the ambiguity arises because the incentives created by the benefit offset vary with the beneficiary's earnings under current law. T1 subjects who would have had no earnings or earnings below BYA under current law are expected, on average, to increase their earnings under the benefit offset. Conversely, some T1 subjects who would have had earnings above BYA under current law are expected to lower their earnings under the benefit offset.[27] For a positive impact on total earnings to occur, the positive impact expected

[26] We calculate total earnings as the total mean annual earnings, as captured in IRS data, across all beneficiaries in our samples.

[27] Empirically, evidence suggests that some high-earning beneficiaries will reduce their earnings but not reduce employment. Weathers and Hemmeter (2011) found evidence of a reduction in earnings by beneficiaries earning above SGA before random assignment in the Benefit Offset Pilot Demonstration.

for those whose earnings would be less than BYA under current law would have to be larger than the expected negative impact on those who would earn more than BYA under current law.

Exhibit 2-2. Definitions of Confirmatory and Exploratory Administrative Outcomes and Predicted Direction of Impacts, if Any

	Definition	Predicted Direction
Confirmatory Outcomes		
Total earnings in 2014	2014 annual earnings	?
Total SSDI benefits paid in 2014	Sum of SSDI benefit payments from January through December 2014; for SSDI workers, it includes benefits for dependent spouses and minor children, but not for a Disabled Adult Child (DAC)[a]; for DAC and a Disabled Widow/Widower Beneficiary (DWB), it includes benefits payable only to the DAC or DWB	?
Exploratory outcomes		
Earnings Outcomes in Calendar 2014[b]		
Any employment in 2014	Indicator for2014 earnings greater than $0	+
Earnings above BYA	Indicator for 2014 earnings greater than or equal to $12,840 (non-blind subjects) or $21,600 (blind subjects)	+
Earnings above 2 times BYA	Indicator for 2014 earnings greater than or equal to $25,680 (non-blind subjects) or $43,200 (blind subjects)	?
Earnings above 3 times BYA	Indicator for 2014 earnings greater than or equal to $38,520 (non-blind subjects) or $64,800 (blind subjects)	?
Benefit Outcomes for January–December 2014		
Number of months with SSDI payments	Number of months with SSDI benefits paid above $0	+
Total SSI benefits paid	Sum of SSI benefit payment amounts from January through December 2014	–
Number of months with SSI payments	Number of months with SSI benefits paid above $0	–

[a]For a description of family benefits, see http://www.socialsecurity.gov/pubs/10024.html#a0=3; accessed May 27, 2014.

[b]Earnings relative to BYA are based on earnings reported in the MEF.

Similarly, the predicted impact of the $1 for $2 benefit offset on SSDI benefits depends on the earnings of the beneficiary under current law. For those who would have had no earnings or earnings below BYA under current law, the offset's predicted impact on benefits is negative, on the expectation that some beneficiaries will earn more than BYA under the offset and hence receive partial benefits. Conversely, for many of those who would have had earnings above BYA under current law, benefits under the offset are expected to be higher because beneficiaries receive partial rather than no benefits, as under current law. Hence, to generate a reduction in mean benefits paid, the reduction in benefits paid to those whose earnings would be less than BYA under current law—but who move above BYA because of the offset incentive—must exceed the increase in benefits paid to those who would earn more than BYA under current law.

Theory predicts positive impacts for three of the exploratory outcomes: the percentage of beneficiaries with employment, the percentage of beneficiaries with earnings above BYA, and months with SSDI payments. It also predicts negative impacts on the earnings of those who would have earnings substantially above BYA under current law; we expect that such individuals would continue to earn above BYA but that they may reduce their earnings because they can maintain the same level of income with fewer hours of work under the offset relative to current law. These predicted reductions in earnings for some beneficiaries mean that theory does not predict an increase in mean earnings over all treatment subjects, despite the predictions of increases in the percentages employed and with earnings above BYA. Finally, theory predicts increases in months with SSDI payments because some beneficiaries who would have had their benefits suspended under current law will receive partial benefits under BOND, even without a reduction in earnings.

Theory predicts negative impacts on SSI benefits and months with SSI payments. Under current law, any beneficiary who concurrently receives SSDI and SSI (a concurrent beneficiary) and is engaged in SGA after completing the TWP and GP is entitled to, at most, only an SSI payment.[28] In contrast, a concurrent T1 subject with the same earnings would likely receive a partial SSI benefit, and the size of the T1 subject's SSI benefit would be reduced by the amount of the partial SSDI benefit or by the entire current-law SSI payment if the latter is smaller than the partial SSDI benefit. The offset might also have an impact on SSI payments to SSI subjects who are SSDI-only beneficiaries at the outset of the demonstration and whose SSDI benefits are below the maximum federal SSI benefit amount. Under current law, some such subjects are likely to enter SSI after they spend down their assets to the point at which they satisfy the SSI resource test. Higher earnings under the offset might reduce or slow the entry of such SSDI-only subjects into SSI and thus reduce SSI payments and months with benefits.[29]

For the two remaining exploratory outcomes—earnings above two times BYA and earnings above three times BYA—theory does not clearly predict the direction of impacts. For those treatment beneficiaries whose earnings would be less than BYA under current law, the offset is likely to have a positive average earnings effect, perhaps increasing the proportions with earnings above two or three times BYA. Conversely, for those who would have had earnings above BYA under current law, the benefit offset is likely to have a negative average earnings effect, perhaps decreasing the proportions with earnings above two or three times BYA. Given that theory does not predict the magnitudes of these opposing expected

[28] Under the SSI Earned Income Exclusion (EIE), monthly SSI benefits are reduced by $1 for every $2 of earnings above an earnings disregard that is as low as $65. Whether a concurrent beneficiary with earnings above SGA is eligible for a federal SSI payment depends on whether the beneficiary's SSI countable income, including earnings not excluded under the EIE and any other countable income, exceeds the maximum federal SSI payment amount. SSI countable income rules exclude $20 of SSDI benefits unless that exclusion is used against some other form of unearned income. Beyond any exclusion, and holding earnings constant, every $1 of SSDI benefits reduces the SSI payment amount by $1 until the SSI payment amount is zero. At any earnings amount above SGA, any SSDI payment under the offset displaces any SSI payment that is due, dollar for dollar. Under BOND, the benefit offset indirectly affects the SSI payment amounts through the SSDI benefit adjustment. For example, for a concurrent T1 subject with earnings above BYA and positive SSI benefit amounts, a $2 increase in earnings would result in a $1 increase in EIE (reducing SSI) and a $1 decrease in SSDI (increasing SSI), which would leave SSI payments unchanged.

[29] See Riley and Rupp (2012).

effects, it not possible to predict the overall direction of impact for either of these higher earnings thresholds.

We did not develop theoretical predictions for the impact of the benefit offset on the frequency and size of overpayments. Before the demonstration, there was no reason to expect that the incidence of overpayments *caused by delays* in adjustments after the beneficiary first completed the TWP and GP would be higher or lower under the benefit offset than under current law. In the absence of an impact on the duration of delays, the switch from adjustment on the basis of monthly earnings under current law to adjustment on the basis of annual earnings under the offset might increase the incidence of overpayments because of late-year changes in earnings that have the effect of retroactively changing the benefit amount due for earlier months of the same year. However, the effect of the offset on the incidence of such overpayments also depends on how the administrative processes for the offset function relative to those under current law. The sign of the expected impact on the size of an overpayment is negative because, under current law, the size of the overpayment in a month with an overpayment is the full monthly benefit; under the offset, however, it may be less than or equal to the monthly benefit.

Administrative features of the offset that could influence impacts on administrative outcomes

The previous discussion abstracts from the administrative features designed and implemented to facilitate use of the benefit offset by T1 beneficiaries. As described in the *Evaluation Analysis Plan*, given that the processes necessarily differ from processes under current law, they are part of the benefit offset—the intervention being tested under BOND.

In the first years of BOND, the administrative features most likely to have affected outcomes concerned the administrative processes leading to the adjustment of benefits—the special processes implemented for T1 subjects and the current-law processes that apply to C1 subjects. For T1 subjects, the process started shortly after random assignments were determined in April 2011. May 2011 was the first month in which beneficiaries could potentially use the benefit offset. Some of those randomly assigned to use the offset informed the demonstration of their work activities, as recommended in demonstration outreach materials. Such beneficiaries eventually saw their benefits adjusted via an administrative process set up for that purpose. It is likely, however, that other T1 subjects who used the offset early in the demonstration did not contact the demonstration. If so, SSA would not have discovered their high earnings until its annual review of earnings reported to the IRS and could only then have initiated the process to adjust benefits.

The benefit measures for the current report are based on benefits paid *in* 2014 rather than on benefits paid *for* 2014, which will eventually include future retroactive adjustments to benefits paid in 2014. These two benefit measures will diverge according to the dollar value of retroactive adjustments made for 2014 benefits. Even though the dollar value of the adjustments is not yet known, we can say with certainty that there will be retroactive adjustments of some dollar amount for the treatment subjects who did not proactively inform SSA of earnings above BYA during 2014. The BOND administrative data as of December 2015 show that SSA did not adjust the benefits of 25 percent of T1 subjects who used the benefit offset by the end of 2014 (i.e., those who had completed their TWP and GP and earned above BYA by the end of 2014) until the following year. In other words, some adjustments to benefits paid to T1 subjects for 2014 are not reflected in benefits paid *in* 2014, and there will be at least some discrepancy between benefits paid *in* 2014 and benefits paid *for* 2014.

The implications of lagged adjustment for impact magnitude depend on how the adjustment processes for the T1 group compare to the corresponding processes for C1 subjects. The most striking difference is that T1 subjects had to be notified about a change in the earnings rules before the benefit adjustment process could start, whereas C1 subjects were subject to rules that had been in place for many years. Further, T1 administrative processes had not previously been implemented on a large scale, resulting in start-up delays.[30]

One other administrative factor that seems likely to have a positive impact on benefits paid *for* 2014, and possibly on benefits paid *in* 2014, is the change from monthly to annual accounting. The aim of annual accounting was to simplify administration of the benefit offset and to simulate the expected future accounting procedures in the event that the offset becomes national policy. The move to an annual accounting period is also expected to assist beneficiaries with highly variable month-to-month earnings (for example, seasonal workers). Under monthly accounting, earnings above SGA in any month reduce benefits for that month, but, under annual accounting, the benefit reduction caused by the same earnings will be smaller or zero because earnings below the SGA amount in other months of the same year keep annual earnings closer to or below BYA. Holding earnings constant, this administrative change is expected to increase the benefits paid to some beneficiaries. Thus, some beneficiaries with variable earnings may have new opportunities to increase their earnings without any reduction in benefits.

Final analysis sample size used to estimate administrative outcomes

The final Stage 1 analysis sample contains a total of 968,713 subjects spread across the T1 (77,115) and C1 (891,598) groups.[31] The Stage 1 analysis sample is nationally representative of SSDI beneficiaries. As would be expected if random assignment were properly implemented, all differences in baseline characteristics between the two groups are small and appear to be attributable to chance. In the *Stage 1 Early Assessment Report*, an omnibus test for differences across all characteristics shows no statistically significant difference between groups. Baseline equivalence increases our confidence that any impact estimate differs from zero at a statistically significant level and represents a real impact of the interventions rather than systematic preexisting differences between the two groups or their environments or chance differences in outcome levels.

Survey sample size and outcomes analyzed

The Evaluation Team completed interviews with 2,916 T1 subjects and 2,819 C1 subjects. Weighted, the survey samples represent all beneficiaries in the population[32]—that is, the same population for which impacts on administrative outcomes are calculated. For more details, refer to Appendix B.

In this report, we present impact estimates for more than 250 outcomes developed from the survey information. Each impact estimate measures the impact of the Stage 1 intervention on T1 subjects,

[30] This issue is described in the *Process Study Report*.

[31] The final Stage 1 analysis sample used for the impact analysis excludes pairs of related beneficiaries who receive disability benefits based on a common primary beneficiary's record if the two members of the pair were randomly assigned to different Stage 1 experimental groups (T1 versus C1). See the *First-Year Snapshot of Earnings and Benefit Impacts for Stage 1* for details.

[32] Subject to the fact that at the time of the survey the T1 and C1 groups could have experienced a different rate of ineligibility.

relative to current law as represented by C1 subjects. We consider all survey outcome impact analyses as exploratory.

We report survey outcomes in three chapters. Chapter 5 has information regarding beneficiary perspectives about BOND whereas Chapter 6 focuses on survey information about beneficiary pathways to employment. Chapter 10 contains information on all other survey-based outcomes, which span several domains, including employment, fringe benefits, employment-related expenses, employer accommodations, income, material hardship, benefit receipt, living situation, health outcomes, health insurance, marital status, and time use.

Relative to the administrative outcome data, the survey data have multiple advantages. One advantage is that they allow for outcomes to be defined over time periods relative to the time of demonstration random assignment rather than periods that are tied to calendar years as is required when using SSA earnings data. A second advantage is that they enable analysis of a broader set of outcomes than are recorded in the administrative data.

Relative to the administrative data, for impact estimation, the main disadvantage of the survey data is that the power to detect impacts is much more limited because of relatively small sample sizes. As described in the *Evaluation Analysis Plan*, the extraordinarily large sizes of the T1 and C1 samples were chosen in order to allow estimation of impacts on outcomes from administrative data. For cost reasons, the survey samples are much smaller than the full T1 and C1 samples, so the power to detect impacts on outcomes measured only in survey data is much smaller as well.

A second disadvantage of the survey data is non-response. It is possible that survey-based impact estimates differ from what they would have been had survey data been available for all subjects selected for the survey. For all estimates presented in the chapter, we used survey weights designed to minimize non-response bias.[33] The overall response rate to the Stage 1 36-Month Survey was 58.6 percent. T1 survey subjects completed the survey at a slightly higher rate (59.1 percent) than C1 survey subjects (58.2 percent).

A third disadvantage of the survey data is that responses may be inaccurate, because of limited recall or concern about the confidentiality of their responses. Because the survey data examined in this chapter are almost all point-in-time measures, recall error is likely inconsequential. Some respondents might have elected to not report or under-report earnings because of concerns about the consequences for their benefits, despite strong assurances that their responses would be used only for research.

Impact estimation methodology for administrative and survey outcomes

The goal of Stage 1 of the demonstration is to make inferences about the impact of the benefit offset if it 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 point estimates and their standard errors for a nationwide population. The standard errors reflect random variation associated with

[33] The non-response weights cause survey respondents whose baseline characteristics are most similar to survey non-respondents to have the greater influence on analytic results, offsetting (in measurable, but not in unmeasurable, ways) the absence of the non-respondents from the analysis sample.

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

To estimate impacts, we compare mean outcomes on a given measure (for example, 2014 earnings) for the T1 group to the mean of the same outcome for the C1 group. For outcomes derived from administrative data, the sample means are weighted for differences in (i) site-selection probabilities and, (ii) sampling rates into T1 and C1 status across sampling strata. For outcomes derived from the Stage 1 36-Month Survey, in addition to the factors for which the administrative data are weighted, the sample means are weighted for differences in survey sampling probabilities within strata defined by 11 categorical measures of beneficiary characteristics from administrative data and for a beneficiary's propensity to respond to the survey once sampled, in order to address the possibility of non-response bias.

For both administrative and survey outcomes, means are adjusted for the effects of small random differences in baseline characteristics.[34] 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 only 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 (i.e., 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 (i.e., 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 (i.e., 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 the applicability of the benefit offset rules to the earnings of *all* T1 subjects, whether or not those subjects work and use the offset. Hence, our average impact measures reflect no impacts on T1 subjects who do not respond to the offset or have their earnings or benefits in any way 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) only those beneficiaries who use the offset.

We make a multiple-comparison adjustment for the two confirmatory outcomes—outcomes selected on the basis of theory and policy interest alone (see the *Evaluation Analysis Plan*). The adjustment is needed because we are testing several outcomes, thereby making the probability of a Type I error (rejecting the null hypotheses if it is true) larger than the significance level for the individual tests. To compensate, we adjust the test statistics for each of the two confirmatory outcomes—2014 earnings and total SSDI

[34] Appendices A and B provide a full description of the estimation method and the construction of analysis weights.

benefits paid *in* 2014—so that the probability of rejecting the null hypothesis of no impact on either confirmatory outcome is equal to the specified significance level if the null hypothesis is true.[35]

We make no multiple-comparison adjustment to the tests for exploratory outcomes. 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.

Impacts on beneficiary subgroups

We also estimate impacts for seven pairs of beneficiary subgroups. For each subgroup pair, we examine the nine outcomes taken from administrative data. We treat all subgroup analyses, including the analyses of subgroup effects for impacts on earnings and SSDI benefits paid, as exploratory. The impact estimation method we use for each subgroup mirrors the impact estimation method we use for the entire sample. 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. The duration subgroups are of interest because 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). 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. Hence, if such impacts exist, we are more likely to observe them in later years 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 beneficiaries. 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

[35] Our approach adjusts the *p*-values for the confirmatory outcomes by using the Westfall-Young stepdown method. Appendix A 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.

worked in 2010 may be able to increase earnings enough to take advantage of the offset more readily than beneficiaries 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 otherwise be concerned that they will lose their Medicaid coverage if they enter or return to the workforce. Commercial or employer-based 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 are more likely to seek employment than study subjects without access to such a program, other things equal.[36] Any consideration of the results of this analysis should note that MBI was not randomly assigned. Thus, access to an MBI program may be correlated with other features of the policy or economic environment that also affect impacts; as a result, differences in impacts for the MBI and non-MBI subgroups could also reflect those factors.

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; therefore, it will be interesting to see whether the earnings and benefits of the two affected groups are more or less sensitive to the introduction of the benefit offset relative to beneficiaries with other impairments.

[36] 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. Background Characteristics of BOND Sites and Subjects

Stage 1 of BOND was designed to produce valid, nationally representative estimates of the impact of the benefit offset for all SSDI beneficiaries. Toward that goal, the 10 randomly selected BOND sites reflect national variation in their environments. Understanding the background of the sites—and the research sample members living in them—provides useful context for interpreting the study findings while informing any future implementation of a national program. Although the evaluation does not estimate site-specific impacts, knowledge of site-level variation in background characteristics and changes in site environments during BOND also contributes to understanding the study findings.

The BOND sites differ in six salient ways: (1) geographic characteristics; (2) nature and strength of the labor market; (3) presence of non-BOND SSDI benefits counseling services; (4) number and staffing configuration of BOND benefits counseling providers; (5) availability of employment services and other work-focused, disability-related resources; and (6) number of BOND T1 subjects. We summarize most of these dimensions in Exhibits 3-1 and 3-2 and discussed them in more detail in the *Process Study Report* and the *Stage 2 Early Assessment Report*. The current chapter—Sections 3.1 to 3.6—provides a summary of this information and, where relevant, describes changes and new observations.

A description of BOND subjects also provides context for the evaluation. In addition, differences between T1 and C1 subjects' baseline characteristics should be small enough that they can be attributed to chance, providing evidence that random assignment was properly implemented. Accordingly, Section 3.7 expands on previous reports from both perspectives—the characteristics of all BOND Stage 1 subjects and with respect to T1/C1 differences—by providing new information from the Stage 1 36-Month Survey on characteristics of Stage 1 subjects that are unavailable in administrative data.

3.1. Geographic Characteristics

Sites vary in the number of states and communities included in their catchment areas, population density, and geographic dispersion of SSDI beneficiaries, as shown in Exhibit 3-1. As discussed in Section 2.2 of the *Process Study Report*, this geographic variation has implications for the demonstration. Service delivery is more complex in sites where providers must understand and navigate multiple sets of state and community policies and resources and tailor service delivery accordingly. For example, the four-state Northern New England site relies on four state vocational rehabilitation agencies (SVRAs) to provide services to beneficiaries. In contrast, Greater Detroit is contained entirely within the state of Michigan. In addition, benefits counseling staff stated that beneficiaries in rural areas may face challenges regarding access to jobs and employment support services.

3.2. Economic Indicators

In two ways, the relative strength of the local economic environment may affect beneficiaries' opportunities to engage in SGA, a necessary step toward using the benefit offset. First, if there are few job openings, individuals with disabilities may experience difficulty in finding employment. Evidence suggests that, while all workers find it more difficult to secure a job during periods of high unemployment, opportunities for individuals with disabilities worsen even more than for others (Livermore et al. 2012). Second, in a weak economy, declines in state revenues often lead to funding cuts for support services for people with disabilities (Johnson et al. 2011). These factors affect employment options for both treatment group and control group members; therefore, we cannot confidently predict the

Exhibit 3-1. Characteristics of BOND Sites

Site	Geographic Characteristics			Number of Stage 1 Treatment Group Subjects[3]		BOND Work Incentives Counseling (WIC) Providers			Centralized Post-Entitlement Process for WIC
	Number of States	Population Density[1]	Geographically Dispersed[2]	T1s Mailed to[4]	T1 Setups[5]	Number	Types	Dispersed Staffing	
Alabama	Single	94 (AL)	X	11,254	2,862	1	• Nonprofit		X
Arizona/ SE California	Multiple (1 full, 1 partial)	56 (AZ) 239 (CA)		7,787	2,179	1	• Nonprofit		X
Colorado/ Wyoming	Multiple (2)	49 (CO) 5.8 (WY)	X	5,549	1,731	1	• Nonprofit	X	X
DC Metro	Multiple (1 full, 3 partial)	9,856 (DC) 203 (VA) 595 (MD) 77 (WV)		4,222	1,346	2	• For-profit • Other[6]		X
Greater Detroit	Partial	175 (MI)		7,930	2,228	1	• Nonprofit		
Greater Houston	Partial	96 (TX)		6,928	1,848	1	• Nonprofit	X	X
Northern New England	Multiple (3 full, 1 partial)	147 (NH) 43 (ME) 839 (MA) 68 (VT)	X	7,808	2,121	4	• Nonprofit • SVRA • University • Medical Center	X (ME, VT)	X
South Florida	Partial	96 (FL)		12,232	3,253	1	• Nonprofit		X
Western New York	Partial	411 (NY)		7,834	2,295	3	• Nonprofit • Advocacy Organization	X	
Wisconsin	Partial	105 (WI)	X	7,892	2,727	5	• Nonprofit • State Health Agency	X	X
U.S. Average or BOND Total	**N/A**	**87**	**N/A**	**79,436**	**22,590**	**N/A**	**N/A**	**N/A**	**N/A**

Sources: Based on BODS, staff interviews, and additional data collection from BOND site visits. N/A = Not applicable

[1] Population density indicates number of individuals per square mile of land in 2010.
[2] Geographic dispersion defined as 20 percent of the SSDI population living outside the Metropolitan Statistical Area (MSA). See Section 2.2 of the *Process Study Report*.
[3] The control group includes 593,824 C1 subjects.
[4] The total number of T1s mailed a letter to inform them about BOND. This count is slightly less (by 4) than the number of subjects randomly assigned to BOND because there of missing addresses.
[5] A beneficiary's record is officially set up when a BOND staff member has explained the BOND reporting requirements and availability of WIC services to him or her. This includes record setups through December 31, 2015.
[6] Association of disability service providers

Exhibit 3-2. Employment Rates in the BOND Sites, 2011 and 2014

Site	State(s) Partially or Totally Included in Site	Employment Rate for People without Disabilities, age 18–64 (%)			Employment Rate for People with Disabilities, age 18–64 (%)		
		2011	2014	Percent Change	2011	2014	Percent Change
Alabama	Alabama	70.2	71.3	1.6	26.5	27.2	2.6
Arizona/SE California	Arizona California	69.9 69.5	72.5 72.2	3.7 3.9	32.8 31.4	32.8 33.3	0.0 6.1
Colorado/Wyoming	Colorado Wyoming	76.3 80.1	79.1 81.0	3.7 1.1	41.4 47.8	41.6 45.2	0.5 -5.4
DC Metro	District of Columbia	71.5	75.5	5.6	30.0	30.3	1.0
Greater Detroit	Michigan	70.2	74.6	6.3	28.9	29.6	2.4
Greater Houston	Texas	73.5	75.3	2.4	36.9	38.0	3.0
Northern New England	Maine Massachusetts New Hampshire Vermont	78.1 76.9 79.5 80.0	79.9 79.0 81.3 80.4	2.3 2.7 2.3 0.5	31.4 31.7 36.8 36.2	32.5 35.5 40.0 36.2	3.5 12.0 8.7 0.0
South Florida	Florida	70.6	73.4	4.0	29.2	30.1	3.1
Western New York	New York	72.1	74.0	2.6	31.3	33.6	7.3
Wisconsin	Wisconsin	78.7	81.1	3.0	38.7	39.8	2.8
Average across 15 Included States[1]	**--**	**74.5**	**76.7**	**3.0**	**34.1**	**35.1**	**3.2**
Entire United States[2]	**--**	**72.8**	**75.1**	**3.2**	**32.6**	**34.1**	**4.6**

Source: American Community Survey. Data for 2014 come from Tables 2.1 and 2.2 of the 2015 Annual Disability Statistics Compendium, based on data from U.S. Census Bureau, 2014 American Community Survey, American FactFinder, Table B18120; http://factfinder2.census.gov; accessed by compendium authors on October 1, 2015. Data for 2011 come from Tables 2.1 and 2.2 of the 2012 Annual Disability Statistics Compendium, based on data from U.S. Census Bureau, 2011 American Community Survey, American FactFinder, Table B18120; http://factfinder2.census.gov; accessed by compendium authors on September 24, 2012.

[1] Unweighted arithmetic average.

[2] Figures include the 50 states, District of Columbia, and Puerto Rico, weighted by relative population size.

direction of the effect of various local economic conditions on demonstration impacts. Some evidence suggests that employment-related interventions have greater impacts when local economic conditions are stronger (for example, Bloom et al. 2003, Greenberg et al. 2003), but there is also evidence for the opposite relationship—that there are greater impacts during periods of weaker economic conditions (Card et al. 2015). Thus, it is plausible that the offset would have a larger impact in a stronger labor market, but there is no guarantee.

The unemployment rate—the number of individuals age 18 to 64 who are not working but are actively looking for work as a share of the labor force (the employed plus the unemployed)—is the conventional indicator of the strength of the local economy. However, for understanding labor market opportunities among people with disabilities, the employment rate—the number of individuals working as a share of the total population age 18 to 64, including those not looking for work—is likely to provide a more useful proxy than the unemployment rate (Burkhauser et al. 2003). The reason is that the employment rate's denominator contains all potential workers, including discouraged workers (those who have stopped looking for work), while the unemployment rate excludes such workers. Many discouraged workers are people with disabilities—they have left the labor force entirely. Given that a large component of the business cycle is characterized by potential workers who become discouraged and no longer seek work, the employment rate tends to fluctuate more than the unemployment rate over the business cycle, providing a more accurate reflection of work engagement levels of the adult population, especially among people with disabilities.

Leading up to BOND enrollment in 2011, the national employment rate among people without disabilities age 18 to 64 had fallen from 75.0 percent in 2007 (before the 2008 recession) to 72.8 percent in 2011, a 2.9 percent decline.[37] For people with disabilities age 18 to 64, the national employment rate had fallen from 36.2 to 32.6 percent, a substantially larger relative decline of 9.9 percent.[38] These changes were mirrored in the rates for the 15 states represented in the 10 BOND sites. In those states, the employment rate for people without disabilities fell from 76.3 percent in 2007 to 74.5 percent in 2011, a 2.4 percent decline. For those with disabilities, the corresponding decline was from 38.1 to 34.9 percent, an 8.4 percent drop.

From 2011 to 2014, the period of the impact analysis for this report, the national employment rate increased for people with and without disabilities: to 75.1 percent for people without disabilities, a 3.2 percent increase, and to 34.1 percent for people with disabilities, a 4.6 percent increase (Exhibit 3-2). The average rates in the states represented in the BOND sites also increased by a similar amount for people without disabilities, to 76.7 percent (a 3.0 percent increase), but by a smaller amount for people with disabilities, to 35.1 percent (a 3.2 percent increase).

[37] See notes to Exhibit 3-2 for sources for 2011. For 2007, data come from Tables 16 and 17 of the 2009 Annual Disability Statistics Compendium, based on data from U.S. Census Bureau, 2007 American Community Survey, American FactFinder, Table B18120; http://factfinder2.census.gov; accessed by compendium authors on April 16, 2009.

[38] As is true for all surveys, there is some sampling error in the Current Population Survey, the source for these employment rates. The sampling error is greater for people with disabilities than for the larger sample of people without disabilities. Because of the sampling error, estimates of changes in employment rates may be lower or higher than the actual change.

During the same period, the change in state-level employment rates among people with disabilities varied across the 15 states in the BOND sites. The employment rate for people with disabilities fell by 5.4 percent in Wyoming (in the Colorado/Wyoming site) and did not change in Arizona and Vermont (in the Northern New England site). The remaining 12 states all experienced increases in the employment rate for people with disabilities, with the highest percentage increases in two of the other Northern New England states (Massachusetts at 12.0 percent and New Hampshire at 8.7 percent) followed by New York (7.3 percent) and California (6.1 percent, part of the Arizona/Southeast California site).[39]

The state-level employment rates for people with disabilities at the end of this report's beneficiary follow-up period, in 2014, varied across the 15 states included in the BOND sites. Seven of the 10 sites included at least one state with an employment rate for people with disabilities lower than the national average. Despite a decrease in the employment rate for people with disabilities over time, Wyoming still had the highest 2014 employment rate among people with disabilities, at 45.2 percent; Alabama experienced the lowest employment rate for the same population, at 27.2 percent.

3.3. Non-BOND SSDI Counseling Services

The WIPA program provides benefits counseling to SSDI beneficiaries who are subject to current law, including the BOND control group. The WIC services provided to the BOND treatment group are intended to be similar in design and content to the WIPA services available to the control group. It is important to examine the SSDI counseling outside BOND to see if WIC services indeed resemble WIPA services. During the demonstration period, WIPA has experienced two important changes that could have implications for the evaluation.

The first change was the suspension of funding for WIPA when authority for the program ended in June 2012. Despite the suspension, most BOND sites maintained some level of counseling services for C1 subjects until SSA reinstated the program in August 2013 (Section 1.2.1 of this report, and the *Process Study Report*).[40] That is, current-law beneficiaries could also learn basic information about SSA rules regarding work from several sources other than WIPA counselors, including SSA field office staff, SVRA counselors in some states, and various staff at private organizations who are familiar with SSDI program rules through experience.

The lapse of WIPA funding did not affect funding for WIC services to T1 subjects or the nature of those services. The changes in funds available to WIPA grantees did, however, lead to WIC (and EWIC) staffing changes (see Section 3.4) because many BOND counseling providers are also WIPA grantees; some of these organizations needed to reconfigure their staffing in response to WIPA changes in ways that also affected WIC staffing.

---

[39] Some variation in state-level estimates both across states and over time is expected because of sampling error.

[40] During this period, call center staff at the Ticket to Work Help Line provided basic information and referral services. In addition, from April 2013 to April 2014, 20 community work incentive coordinators provided telephonic Employment Success Advisor (ESA) services to beneficiaries who were currently employed, had a pending job offer, or were actively seeking employment. The ESA services were similar to those available through the WIPA program. During the 13 months that ESA services were available, the ESA program served 969 beneficiaries throughout the nation (both BOND and non-BOND beneficiaries).

The second change to WIPA occurred in August 2015, when SSA awarded a new round of grants following a competitive application process. The new round of grants instituted 10 changes to the WIPA program, with the goal of providing more targeted, comprehensive, and intensive services with a larger use of remote delivery. Specifically, the changes were intended to:

1. Encourage the use of remote service provision
2. Require counselors to complete continuing education courses annually
3. Distribute funding based on population
4. Expand community outreach
5. Require firewalls between (1) agencies operating as WIPA providers and (2) employment networks (ENs) under SSA's Ticket to Work program and protection and advocacy (P&A) agencies
6. Increase the proportion of WIPA clients who receive intensive services from 80 to 85 percent
7. Encourage WIPA counselors to refer beneficiaries in the early stage of exploring work to a new national help line for information and referral services
8. Establish benchmarks for WIPA performance
9. Require use of SSA-provided software to prepare Benefits Summary and Analysis (BS&A) reports
10. Ensure a minimum 40 percent time commitment in the work allocation of all WIPA counselors

Following consultations with the BOND Implementation Team and Evaluation Team, SSA decided—for several reasons—not to change WIC services in response to these WIPA changes. The first four changes codified or mirrored practices already in place for administering WIC under BOND. For example, in the *Process Study Report*, the Evaluation Team found that, as the demonstration progressed, WIC counselors interacted with beneficiaries more by telephone or e-mail than in person, which is the goal of the first WIPA change above. The fifth change, requiring a firewall between WIPA service providers and Ticket to Work ENs or P&A agencies, also led to analogous separations between these types of organizations and WIC providers because most WIC agencies also provided WIPA services.[41]

The BOND teams and SSA did not expect the remaining changes (numbers 6 through 10) to create a significant impact on counseling delivery for control group subjects as a whole because the changes occurred at a late stage in the demonstration. Any differences would affect only a relatively small part of the evaluation's follow-up period, as August 2015 was more than four years into the six-and-a-half-year

[41] WIC agencies that also provide WIPA services are now subject to firewall requirements between WIPA providers and ENs. However, three WIC providers are active ENs (defined as having at least one ticket assigned as of March 2015) but are not WIPA providers and therefore are not subject to the new EN firewall. One of these WIC providers is in a site with several other providers and serves only a relatively small number of BOND subjects; a second is a subcontractor of the main WIC provider in its site and also serves a portion of BOND subjects. The third provider is the sole WIC provider in its site.

BOND service period for T1 subjects (May 2011 to September 2017). Moreover, some of the changes (such as BS&A preparation, described in Section 4.4) apply primarily to new clients, and relatively few BOND participants were expected to take up WIC services for the first time during the remaining demonstration period. Finally, as mentioned, the demonstration had already maintained WIC services without changes when funding for WIPA services was suspended in 2012. The decision not to make changes to WIC services in response to the 2015 changes to the WIPA program was consistent with this precedent.

3.4. Number and Staffing Configuration of BOND Benefits Counseling Providers

To deliver BOND WIC and EWIC services to treatment subjects, the BOND Implementation Team contracted with local providers already engaged in disability service delivery. Cross-site variation in available providers and geographic coverage areas led to cross-site variation in BOND provider arrangements. As detailed in Exhibit 3-1, arrangements varied with respect to the number of providers in a site, the type of provider organizations (for example, nonprofit agency, SVRA, or educational institution), and the providers' staffing models (dispersed, in which staff allocate a portion of their time to BOND, versus consolidated, in which most staff devote all of their time to BOND).

Differences across sites in provider arrangements affected several aspects of implementation, including (1) providers' ability to accommodate reductions in the number of their full-time equivalent (FTE) positions over the course of the demonstration, (2) the need for coordination and oversight, (3) counselor knowledge of local systems, (4) accessibility of services to beneficiaries, and (5) currency of counselor skills and training. In particular:

- Providers' staffing arrangements and overall size affected their ability to respond to the demonstration's planned reductions in FTEs. Larger providers such as SVRAs had more options for reassigning staff hours to non-BOND work in response to planned reductions in FTEs. Similarly, sites with dispersed staffing structures had greater flexibility to accommodate changes because multiple staff members combined part-time BOND counseling roles with work supported by other funding sources.
- The number of providers and their staffing arrangements affected the proximity and content of services offered to beneficiaries. Sites that covered larger geographic areas, especially more than one state, were more likely to have multiple providers or dispersed staffing structures. Such arrangements placed counselors closer to beneficiaries across the site and retained counselors with knowledge of local resources.
- Relative to sites with fewer providers or more consolidated staffing structures, sites with a larger number of providers and dispersed staffing structures required greater coordination and oversight from the Implementation Team to ensure that providers and staff conducted demonstration activities consistently and as intended.
- Provider and staffing configurations affected counselors' ability to maintain their skills and engage in related training. Staff in sites with fewer providers and more consolidated staffing structures found it easier to consult with their on-site colleagues for support, meet their training obligations, build expertise, and otherwise keep abreast of BOND policies and procedures. These

factors in turn may have affected the quality of post-entitlement work, such as calculating AEEs.[42] A review by the BOND Implementation Team found that, relative to WIC providers with a consolidated staffing model, WIC providers with a dispersed staffing model made more errors in BOND post-entitlement work.

In addition, two recent changes affected most or all BOND counseling providers. The first change for WIC providers was a reduction in FTEs for WIC counseling staff in December 2014. The Implementation Team had planned this reduction in expectation of smaller caseloads of WIC (and EWIC) clients as the demonstration proceeded.[43] Second, to improve the quality of post-entitlement work, the Implementation Team shifted the majority of post-entitlement work to a centralized team in December 2013. Centralization of this work for WIC providers was implemented in Arizona/Southeastern California, Colorado/Wyoming, DC Metro, Greater Houston, Northern New England, South Florida, and Wisconsin in December 2013, and in Alabama in January 2015. We discuss the implications of these changes for WIC services in Chapter 4.

Finally, as mentioned in Section 3.3, changes to the WIPA program led to changes in BOND staffing because many organizations provide both WIPA and BOND services. Specifically, the loss of WIPA funding after June 2012 led to WIC and EWIC staffing changes in 6 of the 10 BOND sites (see Exhibit 2-3 in the *Process Study Report*). According to a member of the BOND Implementation Team, the later reinstatement of WIPA in August 2013 may have helped provider organizations adjust to upcoming reductions in BOND FTE positions. WIPA funding allowed provider organizations to pay for staff time that was no longer reserved for BOND-related duties.

The August 2015 award of a new round of WIPA grants had a more limited effect on BOND counseling providers, resulting in staffing changes related to WIC in just one site. In that site, the sole WIC provider had been a WIPA provider but did not receive a new grant award. As a result, the one counselor providing WIC services left the organization, even though she was working full-time on BOND at the time. The service provider replaced the counselor with the WIC supervisor, who already had significant prior experience as a WIC.

3.5. Availability and Use of Employment Services and Other Work-Focused, Disability-Related Resources

To engage in SGA and use the benefit offset, some beneficiaries require employment services. Providers of these services include SVRAs and other providers acting as ENs under SSA's Ticket to Work program. WIC and EWIC counselors can refer BOND subjects to such providers, just as WIPA counselors do for control group subjects and other SSDI beneficiaries subject to the current law. For example, a counselor might refer a beneficiary in need of career counseling or assistive technology. However, WIC and EWIC

[42] In Section 4.2, we discuss post-entitlement work, which refers to the activities required to facilitate the BOND benefit adjustment process.

[43] As detailed in the Implementation Team's internal planning documents, the team anticipated smaller caseloads over time because of expectations that (1) BOND subjects who took up counseling services earlier in the demonstration would need less support as time elapsed and (2) relatively few beneficiaries would take up counseling services later in the demonstration.

counselors have reported that these resources have not been consistently available because of waiting lists at SVRAs, the small numbers of other local ENs, and variations in the quality of services. The inconsistent availability of services may affect the timing and extent of offset use (Section 2.6 of the *Process Study Report*).[44] These factors are all external to BOND, and are likely experienced similarly by treatment and control subjects. In Section 6.2 of this report, we provide additional counselor and beneficiary perspectives on the use of employment support services, as drawn from the latest rounds of qualitative data collection and the Stage 1 36-Month Survey.

3.6. Number of BOND T1 Subjects

Across the sites, the number of BOND subjects in the Stage 1 treatment group varies. As reported in the *Process Study Report* and the *Stage 1 Early Assessment Report*, the Implementation Team randomly assigned 79,436 beneficiaries to the T1 group and mailed outreach materials to them in batches between May and October 2011.[45] South Florida and Alabama accounted for the largest number of mailings (12,232 and 11,254 mailings, respectively), and the District of Columbia (DC) Metro area accounted for the smallest number of mailing (4,222 mailings).

As of December 31, 2015, records for 22,590 of these subjects had been set up in BODS, meaning that members of the Implementation Team had spoken with these subjects and explained the benefit offset and WIC services to them. South Florida and Alabama also had the largest number of T1 record setups as of December 31, 2015—3,253 and 2,862, respectively—and the DC Metro area had the smallest number, at 1,346. Across all sites, the demonstration had set up the records of 28.4 percent of T1 subjects at that time. Wisconsin had the highest percentage of subjects with record setups (34.6 percent), with the other sites ranging between 25.4 and 31.9 percent. In Exhibit 3-1, we list the numbers of notified T1 subjects and record setups by site as of the end of December 2015.

3.7. Characteristics of Stage 1 Subjects

This section reports two baseline characteristics of both T1 and C1 beneficiaries that have not previously been available. These characteristics are based on the BOND Stage 1 36-Month Survey and are weighted to reflect the national beneficiary population meeting the BOND eligibility criteria in December 2010, just a few months before random assignments were determined in late April 2011. Even though we conducted the survey approximately 36 months after baseline, we have no reason to expect changes in the particular characteristics examined here for individual beneficiaries.

Exhibit 3-3 shows the race/ethnicity and primary language of T1 and C1 beneficiaries. A majority (61 percent) of T1 subjects identify as white (non-Hispanic), 21 percent identify as black or African American (non-Hispanic), 10 percent identify as Hispanic, and the remainder (about 9 percent) identify as another race/ethnicity or did not provide a response. The large majority (93 percent) of T1 subjects speak primarily English at home, about half of the remaining subjects (4 percent) speak primarily Spanish, and

[44] See Honeycutt and Stapleton (2013) for more information on wait times for SSDI beneficiaries at SVRAs and evidence that long SVRA wait times for beneficiaries have a negative impact on their employment and benefit outcomes.

[45] Initially, the Implementation Team randomly assigned 79,991 subjects to the T1 group but later removed over 550 subjects after subsequently learning that they had died before random assignment.

the rest speak another language or did not provide a response. Consistent with expectations if random assignment was performed correctly, these characteristics for the non-response-weighted T1 sample are statistically equivalent to those for the non-response-weighted C1 sample.

Previously, Exhibit 3-1 in the *Stage 1 Early Assessment Report* and Exhibit 2-2 in the *First-Year Snapshot of Earnings and Benefit Impacts for Stage 1* presented information on other demographic and impairment characteristics of Stage 1 beneficiaries at baseline, such as gender, age, primary impairment, and monthly benefit amount. Those tables showed that the T1 and C1 subjects were statistically equivalent with respect to those characteristics as well.

Exhibit 3-3. Baseline Characteristics of T1 and C1 BOND Subjects from Stage 1 36-Month Survey

	T1 Mean	C1 Mean	Difference (percentage points)	P-Value (for difference between groups)
Race/Ethnicity				
White, non-Hispanic	61.4%	59.3%	2.1%	0.226
Black or African American, non-Hispanic	20.6%	19.6%	1.0%	
Hispanic (all races)	9.5%	10.1%	-0.6%	
Alaskan Native or American Indian, non-Hispanic	2.0%	2.5%	-0.5%	
Asian, non-Hispanic	0.6%	1.1%	-0.6%	
Native Hawaiian or Other Pacific Islander, non-Hispanic	0.2%	0.1%	0.1%	
Some other race or multiracial, non-Hispanic	2.9%	4.1%	-1.1%	
Don't know/partial/refused	2.8%	3.3%	-0.5%	
Primary language spoken in home				
English	92.9%	91.4%	1.5%	0.231
Spanish	3.5%	3.8%	-0.3%	
American Sign Language	0.6%	0.4%	0.2%	
Other	1.3%	2.1%	-0.9%	
Don't know/partial/refused	1.7%	2.3%	-0.6%	

Source: BOND Stage 1 36-Month Survey

Notes: Weights reflecting sample selection and survey non-response are used to ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation.

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

*/**/*** Difference is significantly different from zero at the .10/.05/.01 levels, respectively, using a chi-square test that included the survey weights in the testing procedures.

3.8. Summary

This chapter has described the diversity of the employment and service delivery environments in the BOND sites. As discussed in later chapters, this diversity led to variation in implementation practices within BOND. We would expect to see comparable variation in the implementation of a national program similar to BOND, which would take place in equally diverse environments. Indeed, the Implementation Team randomly selected the 10 BOND sites in order to produce nationally representative results.

In addition, several changes in site environments during the demonstration period to date, along with cross-site differences, may help inform the results of the BOND impact evaluation. First, the national employment rate for people with disabilities has improved since 2011, the year in which the demonstration started. As a result, beneficiaries' opportunities to use the benefit offset may be increasing over time—so, too, are potential employment opportunities for control group members, although treatment group members may take advantage of these opportunities to a greater extent because of the offset. Second, the availability and quality of employment support services for BOND beneficiaries remain inconsistent across sites and, for some beneficiaries, pose a significant challenge to working and using the offset. There is no indication that either the economic environment or the availability and quality of service referrals differed for treatment subjects versus control subjects, with one apparently minor exception: the 14-month interruption in funding for WIPA likely had a small impact on the availability of counseling services (including referrals made by counselors) for some control subjects during that period.

4. BOND Benefits Counseling

Benefits counseling is a key component of BOND. The counseling developed for BOND is intended to enable beneficiaries to understand and take advantage of the benefit offset by explaining the effects of earnings on benefits under the offset and referring beneficiaries to employment support programs, such as SVRAs or ENs. BOND includes two types of counseling:

1. Basic Work Incentives Counseling (WIC), which is by design comparable to the counseling available under current law.
2. Enhanced Work Incentives Counseling (EWIC), which adds more intensive services, including counselor outreach to the beneficiary, the development of a detailed employment support plan, and assistance in helping beneficiaries obtain the resources and support they need to find and maintain employment.[46]

This report focuses exclusively on WIC services because Stage 1 subjects are not eligible for EWIC services. Given that WIC staff serve Stage 2 subjects in the T21 treatment group as well as T1 subjects, workload statistics for WICs reflect staff activities with both types of treatment subjects.

This chapter discusses the design of WIC counseling (Section 4.1), WIC counselor caseloads (Section 4.2), benefits counseling services received by T1 subjects (Section 4.3), and BOND post-entitlement services (Section 4.4).

4.1. Design of BOND Counseling for Stage 1 Subjects

The goal of WIC is to provide subjects in specified treatment groups, including T1, with counseling services that have the same overall intensity as services available to SSDI beneficiaries under the status quo. The intent of Stage 1 is to measure the impacts of the benefit offset when implemented with benefits counseling that is adapted for the benefit rules that apply to T1 subjects, but otherwise does not differ from counseling available to all SSDI beneficiaries. Outside BOND, all SSDI beneficiaries are eligible to receive benefits counseling from a WIPA provider. SSA funds 103 WIPA grantees to provide counseling to SSDI beneficiaries about how earnings will affect their SSDI, SSI, and other benefits. WIPA counselors also refer beneficiaries to employment support programs in their communities, such as SVRAs or ENs. The primary objective of WIPA is to equip beneficiaries to make informed choices about work and earnings given SSDI's benefit rules. When those rules change, the counseling must necessarily change as well. With respect to the provision of information, the only intended change under BOND relative to the status quo is the provision of information about the effects of earnings on benefits under the offset rather than under standard SSDI benefit rules—but this change is notable.

In addition to providing information, WIC staff collect work and earnings reports, assist T1 beneficiaries with completing work CDR forms, and occasionally assist T1 subjects in appealing the outcome of work CDR decisions (see Section 7.2.2). Staff at a limited number of sites also provide post-entitlement

[46] For additional details on the design of WIC and EWIC, see Section 5.1 of the *Stage 2 Early Assessment Report* and Section 5.2 of the *Final Design Report*.

services to T1 subjects (see Section 4.4); such services are defined as collecting information from the beneficiary in order to develop an AEE, collecting and reviewing documentation for non-countable income (which SSA deducts from earnings to calculate benefits), and assisting beneficiaries with submitting all of this information to SSA. This involvement with post-entitlement services is a departure from the responsibilities of WIPA counselors that goes beyond adaption to the benefit offset rules; although WIPA counselors often help current-law beneficiaries interact with SSA with respect to similar matters, they do not have a comparable responsibility to collect and review information and complete forms.

4.2. WIC Counselor Caseloads

The design of WIC services has important implications for the potential evolution of WIC counselor caseloads over the course of the demonstration. Subjects enter the WIC caseload when they first contact WIC staff for either information and referral (I&R) or counseling services at any time from May 2011 onward for the T1 subjects or following enrollment into the T21 sample (which occurred between May 2011 and September 2012). After initial contact, beneficiaries remain on the official caseload record for either the remainder of the availability of WIC services (through September 2017) or through the last (60th) month of their BPP, whichever occurs first. The opportunity for subjects to contact WIC staff runs concurrently with this period. Therefore, the total WIC caseload increases over time, as additional treatment subjects make initial contact with WIC providers.

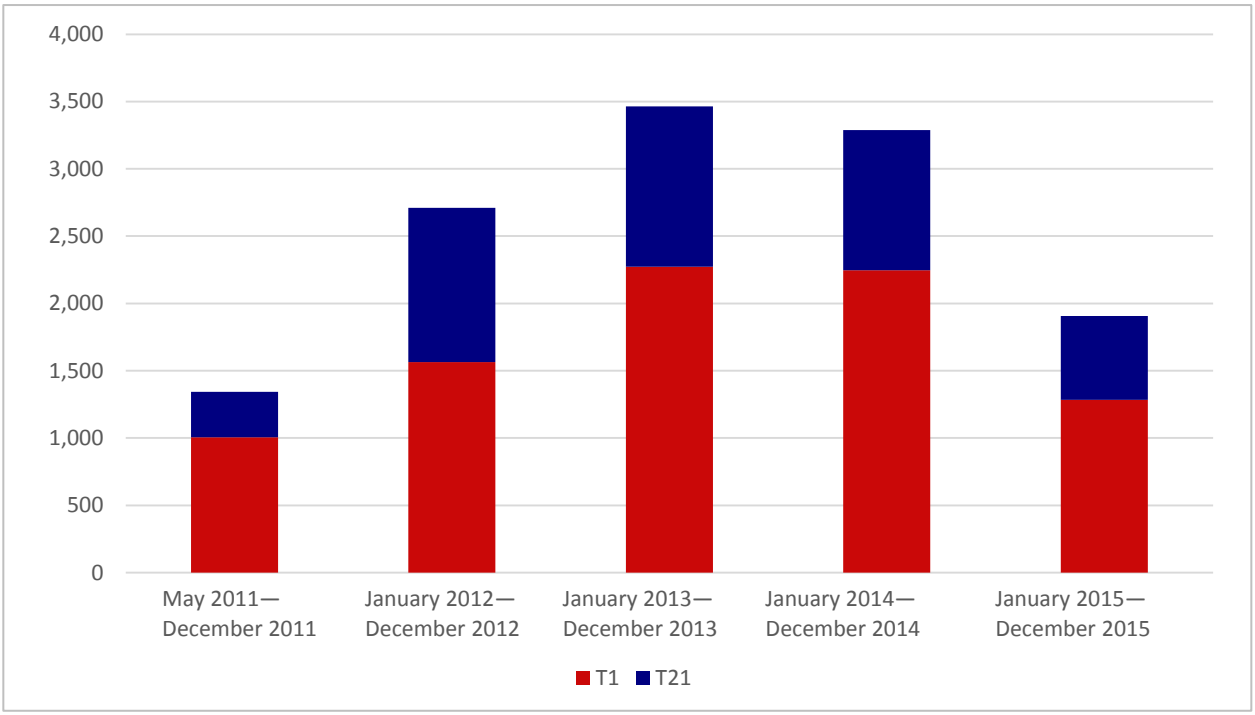
However, not all subjects who enter WIC caseloads continue to actively receive services. Some subjects discontinue contact with their counselors or cycle in and out of service use. Therefore, counselors' actual workloads depend on the number of active cases rather than on total caseload. The Implementation Team expected the active WIC caseloads to eventually start declining. In this section, we define a case as active during a specified time period if the WIC counselor and the subject were in contact or the WIC counselor made a contact attempt.

Exhibit 4-1 shows the number of active WIC cases in each year of the demonstration through 2015. The number of active WIC cases within a calendar year more than doubled between 2011 and 2013, peaking at a high of nearly 3,500 across all BOND sites in 2013. The number of active WIC cases dropped slightly between 2013 and 2014. Between 2014 and 2015, the caseload dropped more sharply to about 2,000 cases in 2015, a decline of about 45 percent. The implications of caseload size and number of active cases depend on the full-time-equivalent (FTE) staffing levels of WIC agencies (i.e., how many counselor hours are available to attend to T1 and T21 subjects' needs), which are described later in this section.

A variety of factors may have affected changes in the number of active cases over time. These factors include the flow of first-time cases into the total WIC caseload, changes in the need for WIC services among subjects already on the caseload due to changes in work activity or personal circumstances, and notices from SSA related to earnings and benefit adjustments. One-on-one telephone interviews with work-oriented T1 subjects in 2015 (described in Chapter 2) provide insights into these factors. Five of 30 work-oriented respondents indicated that, after receiving letters about BOND, they contacted their counselors or other BOND representatives for help understanding the content of the letters or providing information requested in the letters. This suggests that higher 2013 and 2014 caseload activity may in part have been caused by the timing of demonstration implementation activities such as 2011 and 2012 automated reconciliations (conducted in early 2013 and late 2013, respectively), each of which led to

distribution of notices from SSA to T1 subjects related to benefit adjustments, rather than external changes in the subjects' circumstances or interests.

Exhibit 4-1. Active WIC Cases by Year



Source: BTS

Note: The active WIC caseload is defined as the number of T1 and T21 beneficiaries for whom a counselor made a contact attempt or case note in the designated time period. The exhibit includes T21 subjects because T1 and T21 subjects combined make up the total WIC caseload, which is the focus of this subsection. Data from 2011 represent a partial year, as labeled.

Staffing levels at WIC agencies—i.e., the number of FTE positions filled by WIC counselors—affect the agencies' ability to fulfill the needs of the active caseload in each site. The ratio of active WIC clients to FTE slots—the average counselor caseload size—is especially important. As reported in the *Stage 1 Early Assessment Report* and the *Process Study Report*, significant variations in counselor caseload size across sites have existed during the demonstration. Differences continued during the 2014–2015 contract year as summarized in Exhibit 4-2. During that period (December 7, 2014, to December 6, 2015), WIC staff served an average of 150 active beneficiaries per FTE across all sites, ranging from 56 subjects per FTE in the DC Metro site to 289 subjects per FTE in the Wisconsin site. Some of this variation may reflect the Implementation Team's decision to maintain minimum staffing levels at provider organizations serving sites with relatively small caseloads. Site visits also revealed significant variation in counselor caseload sizes among WICs in a particular site, primarily because of differences in counselors' tenure (newer counselors may receive smaller caseloads) and geographic proximity to clients.

Exhibit 4-2. WIC Caseloads per FTE in 2015

	Active WIC Clients (T1 and T21)	FTE Staff	Active Clients per WIC FTE
Alabama	96	1.75	55
Arizona/SE California	266	1.75	152
Colorado/Wyoming	105	1.00	105
DC Metro	56	1.00	56
Greater Detroit	218	1.50	145
Greater Houston	137	1.50	91
Northern New England	242	2.05	118
South Florida	464	2.00	232
Western New York	378	1.48	255
Wisconsin	289	1.00	289
Average	**225**	**1.50**	**150**

Source: BTS.

Note: The active WIC caseload is defined as the number of T1 and T21 beneficiaries for whom a counselor made a contact attempt or case note between December 1, 2014, and November 31, 2015. Staff FTE allocations are based on funding for WIC counselors in the December 7, 2014, to December 6, 2015, WIC agency contract year and exclude supervisory time.

4.3. Benefits Counseling Services Received by T1 Subjects

The previous section described WIC caseloads, which are an indicator of counselors' availability to provide counseling services. This section, drawing on information from BTS on types and dates of services delivered to BOND treatment subjects, describes beneficiaries' experiences receiving BOND counseling services over time.

Exhibit 4-3 displays the percentage of T1 subjects who received any benefits counseling in each of the five demonstration years to date as well as cumulatively, including those who received only I&R. By December 2015, 4.9 percent of T1 beneficiaries had received benefits counseling at some point since study enrollment. Across the years, service receipt rates rise and then decline, consistent with the numbers of active WIC cases in Exhibit 4-1.

It appears that T1 subjects' receipt of counseling under BOND is more common than other SSDI beneficiaries' receipt of counseling under WIPA. Based on published tabulations of non-BOND data, we estimate that fewer than 1.1 percent of SSDI beneficiaries other than treatment subjects meeting BOND eligibility criteria received WIPA services in 2011[47] compared to 1.3 percent of T1 subjects who received

[47] Schimmel et al. (2013) report that WIPA counselors served 59,600 SSDI beneficiaries in 2011. If all 59,600 of these beneficiaries were eligible for BOND in 2011 but not assigned to a BOND treatment group (an estimated 5,455,000), then 1.1 percent of BOND-eligible beneficiaries not in a BOND treatment group received WIPA services in 2011. Many SSDI beneficiaries fall outside the BOND-eligible population, primarily because of BOND age restrictions, and no doubt some WIPA clients did not meet the BOND criteria. Hence, for BOND-eligible, non-treatment beneficiaries, the actual percentage receiving WIPA services is presumably less than 1.1 percent. The estimate of BOND-eligible beneficiaries not in a BOND treatment group assumes that the national

WIC counseling over 8 months of that year (May through December); therefore, the 12-month rate will be even higher, about 1.9 percent if we use a simple ratio adjustment. The moderately higher rate for BOND may reflect several factors, including:

- The effect of active BOND outreach to inform T1 subjects of their new benefit rules (for which there was no equivalent in 2011 for the SSDI caseload as a whole).
- The absence of information about BOND benefit offset rules from other sources (whereas information on benefit rules under current law can be obtained from local SSA offices, SVRAs, other ENs, and community programs).
- The disproportionately large number of BOND beneficiaries on the rolls for fewer than three years at the time of Stage 1 random assignment (because of the planned oversampling of short-duration subjects, who were expected to have a stronger interest in working than beneficiaries as a whole).[48]
- The demonstration's effort to ensure timely availability of counseling for all T1 and T21 subjects (which may have surpassed what WIPA agencies were able to deliver to their service population).

The highest rates of WIC use by T1 subjects took place in 2013 and 2014. During that period, almost 3 percent of T1 subjects received benefits counseling each year. The usage rate declined in 2015, when 1.6 percent of T1 subjects received benefits counseling. Changes in T1 subjects' receipt of benefits counseling over time are consistent with changes in WIC caseload activity described in Section 4.2. As discussed in that section, a variety of factors may influence the number of subjects seeking and receiving WIC counseling services.

BOND-eligible population at the time of random selection was 53/10 times the BOND-eligible beneficiaries in the 10 BOND sites (1,045,840)—or 5,543,005—because the 10 sites were randomly chosen from 53 national sites; site population estimates come from the *Stage 1 Early Assessment Report*. We subtracted the numbers assigned to the three treatment groups from 5,453,005 to estimate the number of BOND-eligible beneficiaries not assigned to a BOND treatment group.

[48] Short-duration subjects were expected to use counseling services more frequently because earlier research has shown that beneficiaries are most likely to complete their TWP during the first five years after award (Liu and Stapleton 2011).

Exhibit 4-3. Percent of T1 Subjects Receiving WIC Services, by Year

Time Period	T1 Subjects who Received any Benefits Counseling in Time Period (%)
Full Demonstration Period to Date (April 2011–December 2015)	4.9
April 2011–December 2011	1.3
January 2012–December 2012	2.0
January 2013–December 2013	2.9
January 2014–December 2014	2.8
January 2015–December 2015	1.6

Source: BTS.

Note: T1 subjects who received counseling are defined as T1 beneficiaries for whom a counselor made a contact attempt or case note during the indicated time period.

Exhibit 4-4 describes the types of counseling services WICs provided to T1 subjects. Once T1 subjects contacted a WIC counselor, they were likely to receive counseling beyond initial I&R services. Among the nearly 5 percent of T1 subjects who received benefits counseling through December 2015, 79 percent received more extensive benefits counseling than I&R. This rate is similar to the 80 percent of WIPA clients whom SSA expected to receive intensive counseling based on the standard in place until August 2015 (Section 3.3). In terms of other services, WIC staff completed a Benefits Summary and Analysis (BS&A) for almost 40 percent of T1 subjects to whom they provided any services during the follow-up period. This one-time activity is designed to give beneficiaries personalized information about how BOND work incentives apply to their individual circumstances and employment goals. WIC staff also referred almost 30 percent of served T1 subjects for employment support services.

In addition, Exhibit 4-4 shows substantial variation across sites in receipt of the different WIC services measured; the patterns are more complex than simply involving higher WIC participation rates in certain sites. For example, no site was consistently in either the top two or bottom two sites among the 10 sites for all the WIC services considered. Service receipt may have been influenced by, among other factors, variation in the WIC services most often needed by the beneficiaries served in different sites, differential availability of appropriate referral options in different communities, and any differences in provider organizations' orientations to providing services of different types. There is no reason to think that this variation differs from the variation that would occur in a national program.

Exhibit 4-4. Percent of T1 Subjects Receiving WIC Services, by Service Type

BOND Site	T1 Subjects Receiving Any WIC Services, 2011–2015 (%) (1)	Of T1 Subjects Receiving Services			
		Received Only Information and Referral (%) (2)	Received Additional WIC Services (%) (3)	Received BS&A (%) (4)	Received Referral (%) (5)
Alabama	3.0	20.2	79.8	29.5	28.9
Arizona/SE California	5.3	27.0	73.0	36.6	22.7
Colorado/Wyoming	4.5	7.7	93.1	53.4	45.3
DC Metro	6.3	35.6	63.6	43.9	14.0
Greater Detroit	4.3	9.2	90.8	47.0	8.6
Greater Houston	4.9	7.9	92.1	17.4	73.2
Northern New England	7.2	24.2	75.8	48.5	26.9
South Florida	4.1	29.8	70.2	17.5	34.1
Western New York	4.3	23.3	76.7	41.0	16.5
Wisconsin	6.7	17.9	82.1	53.2	17.4
All Sites	**4.9**	**20.9**	**79.1**	**38.7**	**28.1**

Source: BTS.

Note: T1 subjects who received counseling are defined as T1 beneficiaries for whom a counselor made a contact attempt or case note between May 2011 and December 2015.

Based on one-on-one telephone interviews with 30 work-oriented T1 subjects in 2015, contact with a benefits counselor appears to have been associated with long-term offset use.[49] All 10 interviewees who used the benefit offset for three or more consecutive years reported contact with a BOND benefits counselor at some point during their demonstration participation; in comparison, 12 of the 20 who did not use the offset or used it only for a short time reported such contact. Of the 22 who reported contact with a benefits counselor (across all interviewees), 80 percent described discussing the advantages of BOND with the counselor. It may be that the beneficiary contacted the counselor only to obtain information about how the offset worked. Although the counselors may have had an influence on offset use beyond providing information, it is not possible to determine whether they did so. Therefore, we cannot tell whether the counselors had a positive impact on offset use.

4.4. BOND Post-Entitlement Services provided by WIC Staff

In BOND, activities conducted with treatment group subjects to collect estimates and documentation of beneficiaries' earnings and submit them to SSA via BODS are called post-entitlement services.[50] Under

[49] The use of the phrase "associated with" is deliberate. We do not give this association a causal interpretation.

[50] More specifically, BOND post-entitlement work includes (1) helping the beneficiary calculate an AEE and (2) documenting and substantiating evidence of non-countable income that should be deducted from earnings to calculate benefits. Non-countable income is used to appeal overpayment decisions from automated reconciliation for past years and is submitted ahead of automated reconciliation so that SSA can use it when conducting the automated reconciliation.

current law, post-entitlement services refer to a broad class of SSA services, including services related to adjustment of benefits because of work—work CDRs and, if warranted, suspension of benefits due to work—usually completed by SSA field office staff. WIPA counselors might advise beneficiaries to obtain such services from a field office, but they do not play a role in the completion of such services. For efficiency reasons, BOND initially tasked WIC and EWIC counselors with helping their treatment clients complete and submit the information required for their benefits to be offset.[51]

In December 2013, centralized staff from the BOND Implementation Team assumed responsibility for providing post-entitlement services in seven BOND sites that needed more support in delivering these services (Section 3.4). The Implementation Team centralized these responsibilities for an eighth site in January 2015. Before centralization, WIC and EWIC counselors in those sites had struggled to provide post-entitlement services, citing challenges in mastering and keeping abreast of procedural changes related to this work (Section 5.4.3 of the *Process Study Report*). After centralization, WIC and EWIC counselors in these sites continued to provide all benefits counseling services for beneficiaries.

In focus group discussions, counselors at sites that switched to centralized post-entitlement work strongly favored centralization because it allowed them to devote more time to benefits counseling. Still, counselors noted some positive aspects of conducting post-entitlement services themselves. For example, some counselors reported that the experience gained by conducting post-entitlement work had helped them become better counselors. Also, in some cases, centralization appears to have led to beneficiary confusion about the separation of tasks between the centralized staff and the counselor. Counselors reported sometimes being unaware of the activities performed by the centralized staff; for example, counselors did not know the time frames for processing forms or mailing letters. Some counselors said that this lack of information contributed to a decline in beneficiary trust.

The Implementation Team has taken steps to promote communication between staff involved in post-entitlement activities and staff involved in work incentives counseling in the centralized sites. WIC and EWIC counselors in all sites now have real-time access to individual BTS records and can see which activities have been completed. The Implementation Team also developed and distributed a manual describing the responsibilities of the two different sets of staff members. In addition, before and during initial implementation of centralization, the Implementation Team held calls with each site's WIC, EWIC, and centralized staff to clarify roles and discuss specific cases.

WIC and EWIC staff at non-centralized sites reported that they spend the majority of their time on post-entitlement work, but that benefits counseling does not go overlooked because they incorporate it into post-entitlement work. Focus group polling showed that approximately two-thirds of counselors in the non-centralized sites favored maintaining control over post-entitlement work because it helps them understand post-entitlement processes and facilitates close relationships with beneficiaries.

[51] The solicitation packet distributed to potential providers of WIC and EWIC services to BOND treatment subjects did not include post-entitlement work in the list of counselor responsibilities. However, the Implementation Team had always planned to have WIC and EWIC staff support post-entitlement work and included these activities in their initial training. The expectation was that counselors would have ongoing contact with the beneficiaries and were therefore well situated to assist beneficiaries with post-entitlement work.

Section 7.2 of this report provides further discussion of this centralization and its impact on the accuracy of beneficiaries' AEEs.

4.5. Summary

To receive WIC services, T1 subjects must proactively contact WIC staff. As of December 2015, nearly 5 percent of T1 subjects had received WIC services at some point during the demonstration. The Implementation Team initially allocated resources to provide WIC services to up to 30 percent of T1 subjects through September 2017. Although WIC services will be available for an additional 21 months, the slowdown in WIC caseloads beginning in 2014 suggests that WIC uptake will remain well below the initially planned WIC capacity. Of the T1 subjects served by a benefits counselor by December 2015, nearly 80 percent received WIC services beyond I&R. This rate is almost exactly the same as the 80 percent of WIPA clients whom SSA expected to receive intensive counseling by the WIPA standard that was in place until August 2015.

WIC service activity has fluctuated over time. Using annual periods, active WIC caseloads reached a peak in 2013 and have since declined. A similar pattern holds for T1 subjects' receipt of benefits counseling over time. Based on a comparison to external data for WIPA service receipt, it appears that T1 subjects received services at a moderately higher rate than did other SSDI beneficiaries.

Across sites, there are variations in the delivery and receipt of benefits counseling in terms of caseloads per FTE, rates of counseling activity, and types of counseling received. These variations may reflect site-specific differences in context, provider organizations, and beneficiary needs, as expected when implementing a large multi-state program.

T1 subjects' experiences with WICs may simultaneously influence and be influenced by subjects' experiences with BOND. A small number of in-depth interviews with T1 subjects suggested that contact with a benefits counselor is associated with long-term offset use, though it is unclear if counselors motivate use of the benefit offset or if offset use leads to contact with counselors. The next chapter explores the potential relationship between beneficiaries' perception and understanding of BOND and their experiences with the program.

5. Beneficiary Perspectives About BOND

This chapter presents information about Stage 1 subjects' knowledge of how benefits and benefit eligibility are affected by working for pay and how T1 subjects describe BOND. The information is based on responses to the Stage 1 36-Month Survey conducted in 2014 (Section 2.1.2) as well as on the in-depth, semi-structured telephone interviews conducted with work-oriented T1 subjects in 2015 (Section 2.1.1).[52]

When interpreting the results presented in this chapter, it is important to keep in mind the differences between the Stage 1 36-Month Survey and the in-depth interviews with work-oriented T1 subjects and the strengths and limitations of each data source. Weighted, the Stage 1 36-Month Survey was designed to generate a representative sample of all T1 and C1 subjects. In contrast, the in-depth interviews targeted a small set of treatment subjects with disability cessation dates (that is, beneficiaries who earned above the SGA threshold after completing the TWP) and varying durations of benefit offset use. The in-depth interviews were designed to develop a fuller understanding of the experiences, perceptions, and behaviors of T1 subjects for whom we expected the offset to be salient. Even though the in-depth interviews provide rich information about 30 work-oriented subjects, that information is not representative of the experiences of all subjects for whom the offset is salient, let alone for the many subjects for whom it is not salient.

Section 5.1 presents survey-based findings on T1 subjects' knowledge about how earnings affect their benefits under the offset and compares that knowledge to C1 subjects' knowledge of how earnings affect their benefits under current law. Section 5.2 then presents findings from the in-depth interviews with work-oriented T1 subjects, and Section 5.3 synthesizes and summarizes the two sets of findings.

5.1. T1 Subjects' Awareness of the Demonstration and Knowledge of How Earnings Affect Calculation of Benefits

We begin this section by reviewing the importance of understanding the benefit offset in the BOND logic model. We then present the survey results for T1 and C1 subjects' knowledge about how earnings affect their benefits. Finally, we present results on how T1 subjects who have heard of BOND describe BOND.

5.1.1. How Understanding of the BOND Benefit Offset Might Affect Employment Behavior as Predicted by the BOND Logic Model

The benefit offset makes work more fiscally attractive for some beneficiaries and therefore might induce them to work more. In order for that to occur, T1 subjects need to understand that the offset makes work more attractive. Results presented here suggest that understanding is incomplete. At a minimum, treatment subjects need to grasp how they can gain from the offer—how they may earn above the SGA level without losing all of their SSDI benefits—in order for them to change their employment behavior in response to the offer. T1 subjects should also understand that the offset applies only after completion of the TWP and GP—the same point at which benefits are suspended under current law for work above the SGA level. This is a complexity that is difficult to grasp. Benefit counseling services are available to help both T1 and C1 subjects. T1 subjects have access to WIC staff with knowledge of the benefit offset and

[52] The overall weighted response rate to the Stage 1 36-Month Survey was 58.6 percent. The weighted response rate for the treatment group (59.1 percent) was almost identical to that for the control group (58.2 percent).

services, whereas C1 subjects have access to WIPA counselors who provide information about benefits under current-law rules (Section 3.3 describes details on WIPA counseling). Furthermore, the BOND Implementation Team provided T1 subjects with information about the offset through the initial random assignment letters, BOND hotlines, and follow-up outreach. For T1 subjects who used the offset, the benefit offset adjustment process itself provided information about the offset. However, the effect of earnings on benefits is likely not salient to a large share of both T1 and C1 subjects because many beneficiaries are uninterested in, or perceive themselves as incapable of, employment. Hence, our expectation before the Stage 1 36-Month Survey was that a large share of both groups would not be well informed about how earnings affect their benefits and benefit eligibility.

The Stage 1 36-Month Survey asked subjects how their benefits and benefit eligibility are affected by earnings above the SGA limit. Except for the 1 percent of survey responses completed by proxies, the survey asked the questions directly of all Stage 1 subjects. Their responses allow us to address the important questions of "How well did study subjects understand the basic details of the offset offer 36 months after random assignment?" and "How did T1 subjects describe the BOND program at that time?" The sections below describe the questions and the responses of the Stage 1 subjects.

5.1.2. Stage 1 Subjects' Understanding of How SGA-Level Earnings Would Affect Benefit Amounts

In the Stage 1 36-Month Survey, interviewers asked all subjects (except those represented by proxies) how their monthly disability cash benefits would change if they were to earn more than the SGA limit after completing the TWP. We expected the responses of the T1 subjects to differ, on average, from those of the C1 subjects because the two groups are subject to different rules (i.e., offset rules versus current-law rules). However, we also expected that many respondents in both groups would not know the correct relationship between earnings and benefits or would respond incorrectly because the effect of earnings on benefits is likely not salient to a large share of both T1 and C1 subjects.

The questions on the survey were:

> Introduction: *"Under the current rules of the Social Security Disability Insurance program, disability beneficiaries are allowed to earn up to $1,070 per month without a change to your benefits. This limit is called the level of Substantial Gainful Activity or SGA and the Social Security Administration increases this limit each year to adjust for inflation. When disability beneficiaries go to work while receiving disability benefits, SSA ignores the cap of $1,070 for up to nine months, no matter how much a beneficiary earns from work."*[53]
>
> *"We'd like to know which of the following things you think would happen to your monthly disability cash benefits if you were to work and earn more than the SGA limit of $1,070 per month after those initial months have passed. Thinking about the amount of your disability cash benefits, if you earned more than $1,070 after those initial months. . ."*

[53] The interviewer provided the respondent with the SGA for beneficiaries who are not blind, even though the SGA for beneficiaries who were blind in 2014 was $1,800. Approximately 2 percent of Stage 1 subjects had a primary impairment of severe visual impairment (Stapleton et al. 2013).

- *Do you think your benefits would stay the same?*
- *Do you think you would lose your monthly benefits completely?*
- *Do you think your benefits would be reduced but that you would be able to keep receiving some of your monthly disability benefits?*
- *How do you think those benefits would be reduced?*

From the responses to these questions, the Evaluation Team classified subjects as demonstrating one of the following five perceptions:

- Benefits would stay the same
- Benefits would be reduced but not to $0
- Benefits would be reduced to $0
- Benefits would neither stay same nor be reduced
- Don't know whether benefits would change

If subjects stated that they thought that their benefits would be reduced but not to $0, they were further classified by their perception of the reduction amount:

- *By the full amount of benefits (equivalent to "reduced to $0")*[54]
- *By half, a $1 reduction for every $2 in earnings*
- *By some other amount*
- *Don't know how much reduction*

The correct response for T1 subjects is that "Benefits would be reduced but not to $0" and that benefits would be reduced "By half, a $1 reduction for every $2 in earnings." The correct response for C1 subjects is that "Benefits would be reduced to $0."

We find that the percentage of T1 subjects providing answers that are correct for the BOND benefit offset is well below the percentage of C1 respondents who provided answers that are correct under current law (Exhibit 5-1). Among all T1 respondents, 29 percent said that their benefits would be reduced if they earned more than the SGA level after completing the TWP and GP, though not to $0.[55] Among all C1 respondents, 54 percent said that, under the same earnings scenario, their benefits would be reduced to $0.

[54] This category is for the few subjects who gave the inconsistent answers that they did not think that they would completely lose their benefits but that they thought their benefits would be reduced by the full amount of their benefit.

[55] The slight difference in the calculated percentage of T1 subjects who provided the correct answer (28.8 percent) and the calculated percentage who said benefits would be reduced but not to $0 (29.0 percent) is due to the regression adjustment for baseline characteristics.

The difference between the percentage of T1 and C1 subjects that provided the correct answer is statistically significant. The 29 percent of T1 respondents providing the correct answer is essentially the same as the percentage of T1 subjects who had been in contact with the demonstration staff by the end of 2014, as indicated by record setups in BODS (28 percent).[56]

A comparison of T1 and C1 responses suggests that the 29 percent of T1 respondents providing the correct answer overstates the percentage of T1 subjects with a correct understanding of the offset. Specifically, 22 percent of C1 subjects provided the same answer, even though they were incorrect for their circumstances. The difference in the percentages for the two groups is a statistically significant 6.6 percentage points. That is, it appears that the percentage of T1 subjects induced to believe that their benefits would be only partially reduced under the benefit offset is less than 7 percent. It is possible that this small difference is attributable to crossover effects on C1 subjects; that is, some C1 respondents who said their benefits would be partially reduced might have been induced to believe, incorrectly, that their earnings were subject to the benefit offset. Such an explanation seems unlikely to be the primary reason for the C1 result because just 5 percent of C1 respondents indicated that they had heard of BOND (second row of Exhibit 5-1). It seems more plausible that a substantial share of the respondents in both groups who said their benefits would be only partially reduced is poorly informed and guessed this answer. This explanation is bolstered by the finding that less than half (44 percent) of T1 respondents who said their benefits would be only partially reduced indicated that they had heard of BOND (not shown in exhibit).

Treatment subjects who did not provide the correct response gave responses that were either consistent with current law or indicated a lack of knowledge about the relationship between earnings and benefit amounts. Over half of T1 subjects (53 percent) gave a response that is correct under current law: benefits would be reduced to $0; it appears that they had no expectation of being able to keep part of their benefits. That percentage is essentially the same as the percentage of C1 subjects that provided the same answer—the correct answer in their case. An additional 15 percent of T1 subjects provided answers indicating that they did not know the relationship between earnings above SGA and benefit amounts, saying that they either did not know how benefits would change or that they believed their benefits would neither stay the same nor be reduced. That percentage is lower than the corresponding percentage for C1 subjects. Overall, the small differences in the share of subjects giving various responses in the T1 and C1 groups are not as large as the differences that would be observed if the percentage of T1 subjects having a correct understanding of how earnings affect benefits under the offset (29 percent) was instead approximately equal to the percentage of C1 subjects with a correct understanding of current law rules (54 percent). In that event, the difference between the percentage of T1 and C1s who believed their benefits would be partially reduced would likely be closer to 30 percentage points instead of 6.6 percentage points.

[56] A beneficiary's record is considered set up when a BOND staff member affirms that both WIC services and BOND reporting requirements were discussed with the beneficiary.

Exhibit 5-1. Estimated Impacts on Stage 1 Subjects' Understanding of How Earnings Affect Benefits

Outcome	T1 Mean	C1 Mean	Impact Estimate
Demonstrated an understanding of benefit adjustment consistent with respondent's earnings' rules (%)	28.8	54.1	-25.2*** (1.9)
Have heard of BOND (%)	35.4	5.0	30.4*** (1.5)
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)
Subjects who think reduction amount would be[a]:			
By the full amount of benefit (equivalent to "reduced to $0") (%)	0.8	**0.9**	-0.1 (0.5)
By half, a $1 reduction for every $2 in earnings (%)	**12.9**	7.8	5.1*** (1.4)
By some other amount (%)	**9.9**	9.2	0.7 (1.5)
Don't know how much reduction (%)	5.4	4.5	0.9 (0.8)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: 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. Weights reflecting sample selection and survey non-response ensure that the BOND subjects who met the analysis criteria are representative of SSDI recipients in the nation. 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). The 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 appear in parentheses.

[a] The four rows of this panel sum to the percentage responding benefits would be reduced but not to $0.

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 (with no multiple-comparisons adjustment).

The bottom panel of Exhibit 5-1 shows how much beneficiaries think their benefits would decline for those T1 and C1 subjects who expected a decline. The wording of the survey questions leaves some ambiguity about the correct response for beneficiaries subject to the offset rules, though we do not think such ambiguity had a substantial effect on the responses.[57] Among the T1 subjects, 23 percent gave one of

[57] The wording of the survey question said that benefits would be reduced "by half the amount of your benefits, that is, a $1 reduction in benefits for every $2 you earn for work." This is technically incorrect, as offset rules reduce benefits $1 for every $2 a beneficiary *earns above the BYA amount* (the annualized version of SGA)

two possible correct responses (that benefits would be reduced by half or by some other amount). We find that 8 percent of C1 subjects mistakenly thought that their benefits would be reduced by half. It is possible that some of the latter group believed that they were subject to BOND offset rules, but, for reasons discussed previously, it seems likely that they were guessing. By design, the Implementation Team did not inform C1 subjects about the BOND benefit offset; therefore, any knowledge the subjects might have had about the offset came from other sources.

We expect beneficiaries to be more informed about the earnings rules if the rules are salient to their situation. We further expect that the rules, on average, would be more salient to those who were employed at baseline than those who were not employed. We test this hypothesis by replicating Exhibit 5-1 for two beneficiary subgroups: those employed in 2010 and those not employed in 2010, according to administrative records (Exhibit 5-2). Indeed, T1 survey respondents who were employed in 2010 were 10 percentage points more likely to have heard of BOND than those who were not employed in 2010. Also as expected, the share of T1 subjects providing the correct answer to how earnings above SGA affect their benefits is larger for those who were employed in 2010 than for those who were not, but the differences are not statistically significant. Although the differences are fairly substantial in magnitude, the probability that such differences would occur by chance is also fairly substantial. Among the subset of T1 subjects who provided the correct answer to how earnings above SGA affect their benefits under the offset, a statistically significant higher proportion of T1 respondents employed in 2010 correctly identified the size of the reduction than did T1 respondents who were not employed in 2010. Employed T1 respondents were also less likely to say that they did not know whether benefits would change than those who were not employed. This provides some evidence that T1 subjects employed in 2010 had a relatively better understanding of the benefit offset rules than did their jobless counterparts. Nonetheless, less than half of T1 respondents who were employed in 2010 had heard of BOND (44 percent) and one-third provided the correct answer to how earnings above SGA affect their benefits.

We also found essentially no differences across the T1 subgroups in the percentage who provided the response that is correct for current law. In addition, we found no significant differences between the knowledge of C1 subjects in the two 2010 employment subgroups.

rather than by half the amount of the benefits (as indicated by the first part of the question) or for the entire earnings amount (as indicated by the second part). The Evaluation Team expected that relatively few subjects would change their answer to "by some other amount" because of this fine-grained detail.

Exhibit 5-2. Estimated Differences in Stage 1 Subjects' Understanding of How Earnings Affect Benefits, by Employment Status in 2010

Outcome	T1			C1		
	Employed Mean	Not Employed Mean	Difference	Employed Mean	Not Employed Mean	Difference
Demonstrated an understanding of benefit adjustment consistent with respondent's earnings rules (%)	33.7	28.0	5.7 (3.8)	54.4	54.0	0.3 (3.3)
Have heard of BOND (%)	43.5	33.4	10.1*** (2.9)	5.4	4.9	0.5 (1.3)
If earnings are above SGA level after TWP months, subjects who think						
Benefits would stay the same (%)	4.3	3.1	1.1 (1.2)	3.4	5.3	-1.9 (1.2)
Benefits would be reduced but not to $0 (%)	**33.7**	**28.0**	5.7 (3.8)	25.2	21.8	3.4 (2.9)
Benefits would be reduced to $0 (%)	50.7	53.6	-2.9 (3.0)	**54.4**	**54.0**	0.3 (3.3)
Benefits would neither stay same nor be reduced (%)	2.8	3.0	-0.2 (1.1)	5.3	3.9	1.4 (1.8)
Don't know whether benefits would change (%)	8.5	12.3	-3.7* (1.7)	11.6	15.0	-3.3 (2.5)
Subjects who think reduction amount would be[a]						
By the full amount of benefit (equivalent to "reduced to $0") (%)	1.3	0.8	0.6 (0.6)	**0.7**	**1.0**	-0.2 (0.4)
By half, a $1 reduction for every $2 in earnings (%)	**17.6**	**12.0**	5.6* (2.5)	11.1	7.0	4.1 (2.4)
By some other amount (%)	**9.9**	**9.7**	0.1 (2.2)	9.4	9.2	0.2 (1.9)
Don't know how much reduction (%)	4.9	5.5	-0.6 (1.8)	4.0	4.6	-0.6 (1.4)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: 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. Weights reflecting sample selection and survey non-response ensure that the BOND subjects in both samples who met the analysis criteria are representative of SSDI recipients in the nation on almost all observed characteristics. There is one discrepancy: the weights 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). The statistics in this table have not been adjusted for differences in baseline characteristics because they are not intended as impact estimates. Differences may reflect differences in the percentage of respondents residing in multi-subject households. Any such difference is small, however, because the percentage of such subjects in the national BOND-eligible population is less than 3 percent. Standard errors appear in parentheses.

[a] The four rows of this panel sum to the percentage responding that benefits would be reduced but not to $0.

Unweighted sample sizes: Employed in 2010 T1 = 1,106, C1 = 962; not employed in 2010 T1 = 1,810, C1 = 1,857

*/**/*** 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 (with no multiple-comparisons adjustment).

5.1.3. Stage 1 Subjects' Understanding of How SGA-Level Earnings Would Affect Benefit Eligibility

Except for the 1 percent of proxy respondents, interviewers asked all Stage 1 survey respondents about how their long-run eligibility for disability benefits would change if they earned above the SGA level after the TWP and GP. Unlike our expectations for the responses to how benefits would change, we do not expect different responses if both T1 and C1 groups correctly understand their situations because both groups would eventually lose their SSDI eligibility. In particular, for T1 subjects, the offset rules apply only for the five-year BPP, and SSA will terminate benefits for those who continue to engage in SGA once that period ends. Similarly, SSA will terminate the benefits of C1 subjects who engage in SGA in the 37th month after completing the TWP, assuming that they have already used up their three GP months.

The survey questions were:

- *Do you think you would remain eligible for disability benefits in the future, no matter how much you earn from work? That is, you would never have to reapply for benefits?*
- *Do you think you would remain eligible for disability benefits for a while, but eventually you would no longer be eligible to receive benefits? That is, do you think eventually you would have to reapply for benefits?*

From the responses to these questions, the Evaluation Team classified subjects as demonstrating one of the following perceptions:

- Would remain eligible for benefits indefinitely (never have to reapply)
- Would remain eligible for a while (would eventually have to reapply)
- Would neither remain eligible indefinitely nor have to reapply
- Don't know about whether they would remain eligible or have to reapply

Exhibit 5-3 shows the responses of T1 and C1 respondents. Even though we see more correct responses about future eligibility than about how benefits change with earnings above the SGA level, a nontrivial number of beneficiaries in both groups is still uncertain or has incorrect perceptions about future eligibility. We find that 58 percent of the C1 subjects gave the expected answer that they would eventually have to reapply for benefits; the percentage of T1 subjects giving the same response is not significantly different from that of the control subjects. Another 20 percent of C1 subjects said that they did not know about future eligibility while the remaining C1 subjects had incorrect perceptions about future eligibility. Again, the percentages for T1 subjects in these categories are not significantly different from those for C1 subjects.

We expected that more of those employed in 2010 would have a correct perception about the effect of earnings on future SSDI eligibility than those not employed. The differences were in the expected direction for T1 subjects, but most were not statistically significant (Appendix Exhibit E-1). The only statistically significant difference was that one-fifth of T1 subjects not employed responded that they did not know about future eligibility versus only 15 percent of those employed. For C1 subjects, responses for those employed in 2010 were essentially the same as for those not employed.

Exhibit 5-3. Estimated Impacts on Stage 1 Subjects' Understanding of How Earnings Affect Future Benefit Eligibility

Outcome	T1 Mean	C1 Mean	Impact Estimate
All Stage 1 Subjects			
Would remain eligible for benefits indefinitely (never have to reapply) (%)	10.6	10.5	0.1 (1.2)
Would remain eligible for a while (eventually would have to reapply) (%)	**59.1**	**58.0**	1.1 (2.0)
Would neither remain eligible indefinitely nor for a while (%)	11.2	11.4	-0.2 (1.3)
Don't know about future eligibility (%)	19.1	20.1	-1.0 (1.7)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey non-response ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. 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). The 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 appear in parentheses. The correct answers expected from T1 and C1 subjects are indicated in bold.

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 (with no multiple-comparisons adjustment).

5.1.4. Description of BOND by Stage 1 Treatment Subjects

The interviewers asked the T1 subjects who had heard of BOND how they would describe "the BOND program" to a friend or relative. The Evaluation Team created response codes and applied up to five codes to each open-ended response. A benefit of the open-ended nature of the question means that respondents could not guess the correct answer from among multiple options. A limitation, however, is that some respondents who were familiar with the offset failed to articulate an answer that demonstrated some understanding of the rules.

Over one-third (35 percent) of T1 subjects had heard of BOND, and Exhibit 5-4 shows how those beneficiaries described the program.[58] The percentages in the first column sums to more than 100 percent because the description provided by a given respondent could be coded into as many as five categories. One-third of these T1 subjects could not describe the benefit offset rules at all. The next most frequent response—provided by 27 percent of T1 subjects—is consistent with the benefit offset rules: "Offset program/allows people to work/make more money/not lose benefits." If these are the only T1 respondents that understand the BOND offset rules, then the percentage that understands the rules is no more than 9.4 percent (35.1 percent times 26.8 percent)–well below the 29 percent of T1 respondents providing the correct answer to how earnings above the SGA level would affect their earnings after the TWP and GP

[58] The slight difference in the percentage of T1 subjects have heard of BOND presented in Exhibit 5-4 (35.1 percent) compared to Exhibit 5-1 (35.4 percent) is due to the regression adjustment for baseline characteristics in Exhibit 5-1.

(Exhibit 5-1). The next two most frequent responses, each provided by roughly 10 percent of treatment subjects, were "good/helpful/would recommend" and "help find employment/return to work/job counseling." Less than 5 percent of T1 subjects provided a negative description such as "not helpful/poor/dislike program" or "don't understand program/confusing/complicated."

Exhibit 5-4. Description of BOND by Stage 1 Treatment Subjects Who Had Heard of BOND

Outcome	T1 Mean	T1 Standard Error
BOND		
Have heard of BOND (%)	35.1	0.9
Description of BOND for those who have heard of BOND		
Good/helpful/would recommend (%)	10.5	0.9
Not helpful/poor/dislike program (%)	2.4	0.5
Promotes higher self-esteem/independence/better quality of life (%)	0.7	2.5
Incentive program/encourages people to work (%)	1.8	4.1
Offset program/allows people to work/make more money/not lose benefits (%)	26.8	1.4
Help to find employment/return to work/job counseling (%)	9.8	9.1
Help with job training/education (%)	1.2	0.3
Determine improvements/services needed/effectiveness of services (for people to return to work) (%)	0.7	2.5
Different levels of assistance/different groups/lottery/randomly chosen (%)	1.9	1.3
Don't understand program/confusing/complicated (%)	2.3	4.6
Don't know/don't remember/nothing/refused (%)	32.8	1.4

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Respondents who had heard of BOND were asked how they would describe BOND to a friend or relative. The open-ended responses were coded into the categories shown in the exhibit. The maximum number of codes that any response contained was three. Weights reflecting sample selection and survey non-response ensure that the BOND subjects in both samples who met analysis criteria are representative of SSDI recipients in the nation on almost all observed characteristics. There is one discrepancy: the weights 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). Any effect of this discrepancy on the reported statistics is small, however, because the percentage of such subjects in the national BOND-eligible population is less than 3 percent. Standard errors appear in parentheses.

Unweighted sample sizes: T1 = 2,916

5.2. Perspectives About BOND from Interviews with Work-Oriented T1 Subjects

In this section, we report findings from interviews with 30 work-oriented T1 beneficiaries in 2015 to illustrate their experiences with BOND. Interviewers followed protocols comprised of open-ended questions in order to engage respondents in conversation on various topics, including respondents' awareness of BOND, interviewer assessments of beneficiaries' knowledge of the benefit offset and its features, self-reported understanding of the offset, suggestions for improving understanding of the offset, and overall impressions of the program. As described in Section 2.1.1, the Evaluation Team designed the 30 in-depth interviews to gather information from 10 respondents from each of three groups: (1) work-oriented T1 subjects who had not used the offset, (2) short-term offset users (subjects who used the offset

in a single year), and (3) long-term offset users (subjects who used the offset during three or more consecutive years). Below, we describe any differences in responses by subgroup. In addition, where relevant, we relate the findings to analyses of the Stage 1 36-Month Survey described earlier in this chapter.

All 30 work-oriented T1 interview respondents had heard of BOND before the interview—compared to 35 percent of T1 survey respondents (see Section 5.1.2). This high level of awareness likely stems from the salience of the benefit offset to this group, whose members were selected because they had cessation dates stemming from work before the interview.[59] Further, given the way that we selected the subjects for the interviews, SSA had adjusted the benefits of two-thirds of the interviewees under the offset at some point, and their high level of awareness may have stemmed from interactions with the demonstration or SSA as related to benefit adjustments.

Interviewers also asked respondents to explain how participation in BOND might affect their benefits, when, and for how long. Interviewers followed this open-ended questioning with conversational probes and, after the interview, assigned a score to reflect the interviewer's assessment of the respondent's understanding of the benefit offset. Interviewers gave respondents one point for each of the following items the respondent mentioned in conversation: (1) the principle of keeping some amount of disability benefits rather than losing them completely after earning over a certain amount, (2) the specific reduction of benefits by $1 for every $2 earned over a certain amount, (3) the approximate BYA, and (4) the five-year duration of the BPP. In addition, interviewers gave an extra point for key terms associated with benefit adjustments, including TWP, cessation date, or BOND annual accounting period. We interpret scores of 0 points to indicate no understanding of BOND and its features, scores of 1 to 2 points to indicate a basic but incomplete understanding, and scores of 3 or higher to represent a more complete understanding.

Seventeen of the 30 work-oriented T1 interview respondents received interviewer-assessed scores indicating a basic but incomplete understanding of the benefit offset, 5 of 30 had a more complete understanding, and 8 of 30 had no understanding at all. A large majority of the respondents with at least some understanding of BOND (18 of 22) mentioned that participants could keep some of their benefits instead of losing them completely after earning more than a certain amount. Only the 5 with a more complete understanding could describe several additional features, such as the specific $1 for $2 offset, the BYA, or the BPP.

Work-oriented T1 interview respondents' knowledge of BOND and its features varied by whether and for how long respondents had used the benefit offset. Interviewer assessments of beneficiaries' understanding of the offset were lowest among respondents who had not used the offset (4 of 10 who had not used the offset demonstrated no understanding), followed by long-term offset users (3 of 10 long-term offset users demonstrated no understanding), and short-term offset users (1 of 10 demonstrated no understanding). Among the four respondents who had not used the offset and demonstrated no understanding of it, two recalled receiving letters about BOND but found them confusing. In addition, one of the two who found

[59] The 30 work-oriented respondents were randomly selected from a list of T1 subjects with cessation dates (Chapter 2). In that larger list, cessation dates ranged from four months before the interview to more than seven years before the interview.

the letters confusing was so fearful of losing her benefit that she was afraid to look for work. Another of these four respondents felt that he did not need to understand the offset because he relied on a family member (not a representative payee) to handle his benefits. The finding that three of the long-term offset users interviewed had scores indicating no understanding of BOND is somewhat unexpected, but is largely explained by the fact that two of the three long-term offset users were earning too much at the time of the interview to receive any benefits even with the offset (their benefits had been fully offset to $0 by their earnings). As one explained, she understood BOND offset rules when she was just starting to return to work, but she no longer felt the need to retain the details of BOND offset rules once her disability benefit stopped as a consequence of high earnings.

Respondents' self-reported understanding of the benefit offset rules seems to have been influenced by their perceived need to understand the BOND rules, their use of the offset, and related exposure to BOND. Overall, half of respondents (15 of 30) rated themselves as comfortable with their own understanding of BOND, 30 percent (9 of 30) were neither comfortable nor uncomfortable, and 20 percent (6 of 30) were not comfortable with their understanding of the offset. One respondent who had not used the offset said that he was comfortable with his understanding of the offset even though he did not understand the rules and would learn about the program if he got a job. Two respondents who said that they were neither comfortable nor uncomfortable with their understanding of the offset indicated similar plans. In contrast to these examples, self-described comfort of one's understanding of the offset was lowest among those who had not used the offset (3 of 10 said that they were uncomfortable with their understanding) and highest among short-term offset users (8 of 10 said that they were comfortable).

Among all work-oriented T1 interviewees, 9 provided suggestions for improving beneficiary understanding of the benefit offset. The most common suggestion (4 of 9) was to make letters and pamphlets from SSA easier to understand by using "layman's terms" and simple language. Two of the respondents who offered this suggestion were among the 20 percent (6 of 30) of interviewed T1 subjects who said they had not spoken with a BOND benefits counselor and therefore were likely relying on written materials from SSA to learn about the offset.

In terms of their broader impressions of the benefit offset, about half of all work-oriented T1 interview respondents (16 of 30) stated that the opportunity to use the offset was positive for them in that it allowed them to keep more money from benefits than they could have otherwise and that it was a safety net that helped make it feel less risky to return to work. These positive impressions were most common among long-term offset users (9 of 10) and those who had not used the offset (5 of 10) relative to those who used it for a short time (2 of 10).

5.3. Summary

The survey results presented in this chapter show that most T1 subjects lacked a correct understanding of the BOND benefit offset rules regarding how earnings relate to SSDI benefits. We found that 29 percent of the T1 respondents to the Stage 1 36-Month Survey gave the correct answer to a question about how earnings above the SGA level affect their benefits under the offset rules. In contrast, 54 percent of C1 respondents provided correct answers consistent with current-law rules. Given that current law is implemented nationally, the C1 figure is at least arguably an indicator of the percentage of beneficiaries that would have a minimal understanding of the offset under a national program (assuming that information sources would be comparable to those available today). A comparison of T1 and C1 survey responses suggests that the percentage of T1 subjects who truly had some understanding of the offset at

the time of the survey is less than 29 percent. Further, some evidence from the survey suggest that the percentage of T1 respondents with a correct understanding of the offset is lower: 22 percent of C1 respondents provided the same answer, even though it was the wrong answer for them; and only 9 percent of T1 respondents had both heard of BOND and described it as an offset program that allows them to work, earn more money, and/or not lose their benefits.

The finding would be of less concern if it simply reflected the fact that the benefit offset is not salient to a large majority of T1 subjects. To assess that possibility, we compared responses for those T1 subjects with earnings in 2010 to responses for those with no earnings in 2010. Presumably, the earnings rules are more salient to the former group. The difference in the percentage of employed T1 subjects providing the correct answer is positive but it is not statistically significant; furthermore, only one in three respondents provided the correct answer to how earnings above SGA affect their benefits.

It is important to note that many T1 subjects may have had some understanding of the offset after the completion of initial and follow-up outreach but had forgotten by the time of the survey. The Implementation Team and SSA sent letters to T1 subjects notifying them of their assignment to BOND and informing them of the BOND rules between May and October 2011. Between 2012 and early 2014, the Implementation Team conducted additional outreach by letter and telephone to increase awareness of BOND. The survey was conducted approximately three to three and a half years after the initial outreach, between May 2014 and February 2015. The duration from the follow-up outreach to the survey interview varied widely across respondents—from just a few months to almost three years—because of the duration of the follow-up outreach period. In addition, beneficiaries who were in contact with the demonstration before the follow-up outreach efforts began did not receive additional outreach. Those who had an understanding of the offset at the time of the survey might be largely limited to those who actively pursued use of the offset after the completion of outreach.

We selected work-oriented beneficiaries as interviewees because of the anticipated salience of the offset to their circumstances; all had engaged in SGA after completion of the TWP and GP at some point, and SSA had adjusted the benefits of two-thirds of the subjects under the offset. The interviews revealed that most respondents had at least some understanding of the offset, and a majority understood that they could keep some of their benefits when they had earnings above the SGA level, but a quarter had no understanding at all, and only a small number had an understanding that was substantially complete.

The findings reported here raise substantial cause for concern that the impacts of the BOND benefit offset on earnings, benefits, and other outcomes may be muted relative to what they would be under a national program. Knowledge of the change in the earnings rules is a crucial prerequisite to the impact of the change on all outcomes. Contrary to the intent of the BOND design, it seems likely that T1 subjects' understanding of the benefit offset is substantially lower than what beneficiary understanding would be under a national program at the time of the survey. It may be that T1 subjects for whom the offset is salient have a better understanding of the offset when it matters—when they are making decisions about work—than when they were surveyed. That seems less likely to be true, however, in the period after the end of the demonstration's outreach activities.

6. Employment Barriers and Facilitators

This chapter presents information on how Stage 1 subjects describe the employment barriers they face and their use of employment supports, education, and training in the 36 months after random assignment. The information is based on responses to the Stage 1 36-Month Survey conducted in 2014 (Section 2.1.2), focus groups with WIC and EWIC staff conducted in 2014 (Section 2.1.1), and the in-depth, semi-structured interviews conducted with work-oriented T1 subjects in 2015.

The information presented in this chapter is based on two data sources, each with different strengths and limitations. The first source, the Stage 1 36-Month Survey, is designed to generate a representative sample of all T1 and C1 subjects. As discussed in Chapters 2 and 5, the second source, the small in-depth interview sample, is intended only to provide richer information about the knowledge and experiences of work-oriented subjects and is not intended to be representative of all T1 subjects or even all those who are work-oriented. It is important to keep these differences in mind when interpreting results.

Section 6.1 presents findings on employment barriers faced by Stage 1 subjects, and Section 6.2 discusses findings on the receipt of employment supports. Section 6.3 summarizes and synthesizes the two sets of findings.

6.1. Barriers to Employment

This section describes the responses given by Stage 1 subjects to questions about factors (barriers) that limit their ability to work. As part of the Stage 1 36-Month Survey, interviewers read a list of potential barriers to employment and asked if respondents strongly agreed, agreed, neither agreed nor disagreed, disagreed, or strongly disagreed that each barrier limited their ability to work. Exhibit 6-1 shows the percentages of Stage 1 subjects that either strongly agreed or agreed that certain factors limited their ability to work. Exhibit 6-1 also presents the estimated impact of the benefit offset on subjects' reporting of these employment barriers and whether the impact is statistically significant. Owing to the large number of hypothesis tests in this chapter, it is possible that we report some statistically significant results that are solely attributable to chance, i.e., false positives. We cannot know which findings are spurious in this way and therefore remind the reader that the results are suggestive, not conclusive.

Exhibit 6-1. Estimated Impacts on Barriers to Employment of Stage 1 Subjects at the Time of Survey

Outcome	Treatment Mean	Control Mean	Impact Estimate
Agree or strongly agree that "I am limited in my ability to work because of a physical or mental condition" (%)	88.5	89.9	-1.4 (1.1)
Agree or strongly agree that "I am limited in my ability to work because I do not have reliable transportation to and from work" (%)	27.6	25.9	1.6 (1.8)
Agree or strongly agree that "I am limited in my ability to work because I am caring for children or others" (%)	13.3	13.5	-0.3 (1.4)
Agree or strongly agree that "It is difficult for me to work because I am afraid I will lose my disability benefits" (%)	27.5	23.7	3.7* (1.7)
Agree or strongly agree that "I am limited in my ability to work because I am finishing a school or training program" (%)	4.3	4.9	-0.6 (0.8)
Agree or strongly agree that "Many workplaces are not accessible to people with my disability" (%)	48.6	51.1	-2.5 (2.1)
Agree or strongly agree that "I don't have the skills or training I need to return to work" (%)	37.7	34.2	3.5 (2.3)
Agree or strongly agree that "It will be difficult to requalify for Social Security disability benefits in the future if I work" (%)	36.2	33.4	2.7 (2.0)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

For the most part, T1 and C1 subjects were similar in reported barriers to employment. The only statistically significant difference is that treatment subjects were slightly more likely than controls to agree or strongly agree that it is difficult to work for fear of losing their disability benefits (28 versus 24 percent). A possible explanation for this difference is that T1 subjects had been made more aware of the potential for benefit loss due to earnings by the outreach they received from the demonstration. Exhibit 5-1 presents further evidence of greater concern among treatment group subjects about losing disability benefits due to work; 82 percent of T1 subjects versus 77 percent of C1 subjects thought their benefits would be reduced either to $0 or by another amount if their earnings exceeded the SGA level.

The Evaluation Team asked open-ended questions during in-depth interviews with work-oriented T1 subjects to gather information about respondents' barriers to employment.[60] Interviewers asked respondents to identify the main factors influencing whether and how much they worked and earned. Interviewers then asked respondents the extent to which potential benefit adjustments influenced their decisions about work.

[60] We classify beneficiaries with a SGA-related disability cessation as work-oriented (Section 2.1.1).

Although almost 90 percent of the T1 subjects who responded to the Stage 1 36-Month Survey agreed that physical or mental health conditions limited their capacity to work (see Exhibit 6-1), we would expect the proportion to be lower among the work-oriented T1 subjects who participated in the in-depth interviews, given that all of the latter had worked to the point at which SSA determined a disability cessation date. Indeed, half of the 30 work-oriented T1 subjects interviewed by the Evaluation Team identified issues with their physical or mental health as the main factors determining whether and how much they worked and earned. Some of these respondents said that they worked more in times of relatively good health and less in times of relatively poor health; others described their health issues as a constant limitation on how much they could work. One respondent explained that she recently had surgery and would not be able to resume work until she recovered. Others explained that the symptoms of an ongoing medical condition or frequent, regularly scheduled medical appointments limited their ability to work at all or to increase work beyond part-time employment.

We interviewed three subgroups of work-oriented T1 subjects based on whether and for how long they had used the benefit offset: (1) subjects who had not used the offset, (2) short-term offset users (subjects who used the offset in a single year), and (3) long-term offset users (subjects who used the offset during three or more consecutive years). Among these subgroups, the majority of short-term offset users and those who had not used the offset (7 of 10 for each group) identified issues with their physical or mental health as the main factor determining whether and how much they worked and earned as compared to just 1 of 10 long-term offset users. Instead, long-term offset users were more likely to identify facilitators, such as work accommodations (3 of 10) or personal motivation to work and earn more (3 of 10), rather than barriers to work as the main factor influencing whether and how much they worked and earned.

Barriers to work for an individual can be dynamic. A barrier to work today may not be a barrier to work tomorrow, and vice versa. Some work-oriented T1 interviewees provided examples of how barriers to work can shift. Among interviewees who described their physical or mental health as the primary factor determining whether and how much they worked and earned, some said that, if their health improved, other factors would become the main barrier to working or increasing work. One respondent explained that first he had to be healthy enough to work but that, once working, he could not afford to lose his benefits; otherwise he would not be able to pay his bills during periods of illness. Hence, the fact that a survey respondent indicated health as the reason that he or she was not working does not rule out substantial work or other barriers in the future.

In a separate question, interviewers asked work-oriented T1 subjects to explain how important the prospect of a benefit adjustment was in their decisions about how much to work.[61] Twelve interviewees (40 percent) indicated that the prospect of a benefit adjustment was an important or somewhat important limiting factor in their decisions about how much to work. By comparison, 28 percent of T1 survey respondents expressed reservations about losing their disability benefits due to work (Exhibit 6-1). The difference in responses might reflect the fact that two-thirds of the in-depth interviewees were selected because they had cessation dates but had not used the benefit offset or had used it for a short time and stopped. Those respondents' underlying concerns about losing benefits due to work may partly explain why they had not used the offset or used it only for a short time.

[61] "Benefit adjustment" is the neutral term interviewers used to describe benefit reductions.

Not all interviewees viewed the potential of a benefit adjustment as a reason to limit work and earnings. Two long-term offset users said that benefit adjustments motivated them to work or earn more. Both viewed the opportunity to increase their total earnings as compensating for their benefit reductions. One explained that the expected benefit reduction influenced her to increase her earnings because BOND "was going to let her keep more of her money," which would help her pay her medical bills.

6.2. Service Utilization, Education, and Training

This section presents estimates of the impact of the benefit offset on beneficiaries' receipt of employment supports, education, and training as reported by the beneficiaries in the Stage 1 36-Month Survey. Interviewers asked both T1 and C1 subjects about the types of services or supports that they received to improve their ability to work in the 36 months since random assignment as well as whether they were enrolled in school or training during that period or at the time of the survey. Examples of types of services or supports included any form of transportation assistance, assistive devices, training to learn a new job or skill, or on-the-job training, coaching, or support services. For Stage 1 subjects who had received training to learn a new job or skill or had received on-the-job training, coaching, or support services, interviewers asked about the associated service locations and referrals.

There is little evidence that the offset had an effect on the reported use of employment supports or receipt of education or training. Exhibit 6-2 displays the percentages of T1 and C1 subjects reporting that they received any type of employment support, schooling, or training for a new job or skill in the 36 months since random assignment. The exhibit also shows the percentage of beneficiaries reporting that they needed but did not receive any of these employment supports in the 36 months since random assignment, along with the percentage enrolled in school or taking classes at the time of the survey.[62] None of the differences between T1 and C1 subjects in the exhibit is statistically significant.

[62] Appendix Exhibits E-2, E-3, and E-4, respectively, present details on the types of employment supports received by T1 and C1 subjects, the unmet need for these employment supports, and the types of schooling or training programs received during the 36 months since random assignment. Appendix Exhibit E-5 shows details on the type of schooling or classes in which T1 and C1 subjects were enrolled at the time of the survey. The only statistically significant results show that T1 subjects were slightly less likely than C1 subjects to receive help in keeping a job and to have needed but not to have received transportation assistance.

Exhibit 6-2. Estimated Impacts on Employment Supports, Education, and Training Received by Stage 1 Subjects

Outcome	Treatment Mean	Control Mean	Impact Estimate
Received any type of employment support (%)	53.2	52.5	0.7 (1.9)
Needed any type of employment support but received none (%)	37.3	37.7	-0.3 (2.0)
Received any schooling or training (%)	6.0	5.6	0.4 (0.8)
Currently enrolled in school or taking classes (%)	3.2	3.3	-0.1 (0.8)

Source: BOND Stage 1 36-Month Survey.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

We also did not find any statistically significant differences between T1 and C1 subjects in locations for two types of employment supports: (1) training to learn a new job or skill and (2) on-the-job training, job coaching, or support services (Appendix Exhibits E-6 and E-7). There were statistically significant difference between T1 and C1 subjects in referral sources: T1 subjects were slightly less likely to have been referred by a job coach to both employment supports.

Another perspective on the receipt of employment services comes from the in-depth interviews with 30 work-oriented T1 subjects. Interviewers asked respondents whether they had used employment supports or services during the last several years.[63] Less than a third of interviewees (9 of 30) reported receiving employment services during that period. This figure is somewhat surprising given that the Stage 1 36-Month Survey found that 52 percent of all T1 respondents used at least one type of employment service; moreover, the interview subjects were selected because of evidence of SGA-level employment. It is possible that the work-oriented interviewees—all of whom had worked enough to have disability cessation dates—did not need services in order to work, although we did not systematically explore this question. Four of the 21 respondents who did not receive services did volunteer that they did not need them. For example, one respondent who was not interested in employment supports said that she was able to find a job on her own. Two others expressed an interest in services but were unable to access them.

WIC and EWIC staff provide yet another perspective on the availability of, referrals to, and use of employment services. In our focus group discussions with WIC and EWIC staff in 2014, 43 percent reported that it was difficult or very difficult to connect beneficiaries with employment services, while 35 percent described service access as very or somewhat easy. These figures provide additional evidence that

[63] As part of asking this question, interviewers mentioned that employment services could be services to which they were referred by a BOND counselor, services through vocational rehabilitation or Ticket to Work, or similar services.

service availability is mixed—possibly depending on the site—and that some beneficiaries have greater access to employment services, compared to other beneficiaries. In an online poll,[64] nearly half of supervisors (47 percent) reported a decrease in non-BOND funding of disability-related services in the past year. Most staff (81 percent) reported that they usually refer beneficiaries to SVRAs, but 68 percent also said that they refer beneficiaries to other ENs on occasion. In discussion, participants explained that ENs were not a primary referral source because of a limited local presence and because few beneficiaries were ready to work full-time—a goal many ENs require of their clients.

Both work-oriented T1 subjects and counselors had mixed reactions about whether available employment-related services met beneficiaries' needs. When considering beneficiaries who had received services, 38 percent of WIC/EWIC focus group respondents said that these services met beneficiaries' needs well or very well, but 35 percent reported that services did not meet the needs of BOND beneficiaries (i.e., they reported that services did so "poorly" or "very poorly"). Similarly, more than half of the T1 in-depth interview respondents who received employment support services (five of nine) described the services they received as unhelpful. For example, one respondent attended a resume class but stated that the class was "useless" and did not help him find a job. Another respondent said that the services he received were not designed to help people in his line of work (accounting), making them unhelpful. At the same time, some interviewees (three of the nine who received services) said that the services they received were helpful. For example, one subject used SVRA and other EN services and described both as helpful.

The interviews with work-oriented T1 subjects did not indicate a clear relationship between the receipt of employment services and benefit offset use. One possible explanation is that employment support services were not closely tied to offset use; it is also possible that work-oriented T1 subjects who entered the benefit offset generally did not need employment support services and therefore did not seek them out. Another possibility is that the employment supports received by respondents were not sufficient to help them use the benefit offset. It bears repeating that the interviewees are not representative of all T1 subjects or even all those who are work-oriented beneficiaries.

6.3. Summary

The Stage 1 36-Month Survey results presented in this chapter show that T1 and C1 subjects had similar perceptions about the employment barriers they faced at the time they were interviewed. The only statistically significant difference is that treatment subjects were slightly more likely to agree or strongly agree that it was difficult to work for fear of losing their disability benefits. One possible explanation for this difference is that BOND outreach to T1 subjects heightened their awareness of the potential for at least some benefit loss due to work.

Among the 30 work-oriented T1 interviewees, half identified issues with their physical or mental health as the main factor determining whether and how much they worked and earned. Some interviewees expected improved physical or mental health that would allow them to initiate or increase work; others

[64] Participants were invited (but not required) to respond to an online, multiple-choice poll during the focus group. For the question directed only to supervisors, 17 of 19 responded. For the remaining questions, the number of WIC and EWIC counselors and supervisors who responded to the poll ranged from 37 to 43 out of 51 possible respondents.

described medical issues that limited their work efforts on an ongoing basis. In addition, 40 percent of the work-oriented interviewees described the prospect of benefit reductions as an important or somewhat important limiting factor in their decisions about how much to work.

There is almost no evidence from the Stage 1 36-Month Survey that the benefit offset had an impact on the receipt of employment supports, education, or training. Of the work-oriented respondents who participated in the in-depth interviews (all of whom were in the treatment group), about one-third had received employment support services, but their responses were not indicative of a clear relationship between receipt of employment services and benefit offset use.

7. Offset Use and the Pathway to Benefit Offset Adjustment

7.1. Introduction

Timely and accurate benefit adjustment according to benefit offset rules relies on multiple complex processes. This chapter analyzes how the benefit adjustment processes have functioned in BOND. It also reports the number of beneficiaries who had SSDI benefits adjusted according to offset rules as of December 2015.

The discussion begins with a summary of SSDI benefit rules under current law, describes the benefit offset rules for BOND treatment subjects, and then reports rates of offset use among T1 subjects. Chapter 2 of the *Final Design Report* provides a detailed review of current-law SSA rules and work incentives. The *Stage 2 Early Assessment Report* (sections 6.1 and 6.2) and the *Process Study Report* (section 5.2) describe in detail how the benefit offset works and how it differs from work incentives in current law.

The following current-law SSDI rules and procedures govern both BOND treatment and control subjects as they work:

- During the ***Trial Work Period (TWP)***, beneficiaries are entitled to attempt work without affecting benefits. In 2014, a TWP month was any month in which an SSDI beneficiary had earnings of at least $770 or worked at least 80 self-employed hours. The TWP consists of nine such months in a rolling 60-month window.
- 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 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.

All SSDI beneficiaries are required to report earnings to SSA, which facilitates these steps. Under current law, SSA suspends SSDI benefits in any month in which a beneficiary engages in SGA, during a three-year period referred to as the Extended Period of Eligibility (EPE). Engagement in SGA after the end of the EPE results in benefit termination.

7.1.1. Benefit Adjustment Processes Developed for BOND

After completing the TWP and GP, BOND benefit offset rules differ from current law. Processes used in BOND to adjust SSDI benefits according to the benefit offset differ with whether or not the beneficiary reports earnings as required by SSA. "Front door" entry into the offset refers to what occurs when treatment beneficiaries report earnings as required and otherwise engage with demonstration staff to complete necessary forms. "Back door" entry occurs when—contrary to requirements—beneficiaries do not report earnings and SSA discovers their earnings via other means.

There are four milestones on the administrative path to the first adjustment of benefits under the offset rules for T1 subjects who report work activity and earnings amounts to BOND or SSA staff and enter by the front door:

1. *Sustained earnings sufficient for offset use*: In order to receive a benefit adjustment through the offset, T1 subjects must have sufficient sustained earnings to complete the TWP and GP followed by calendar-year earnings that exceed the BOND Yearly Amount (BYA). When these sustained earnings are achieved, we refer to the year when earnings exceed BYA as a year with "offset use", even though the actual adjustment of a monthly benefit may not occur until late in the year or retroactively after the year has ended.[65]
2. *Work CDR completion*: SSA must complete a work CDR to verify that the treatment subject completed the TWP and GP and to establish when this occurred.
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.
4. *First benefit adjustment*: SSA's BOND Stand Alone System (BSAS), uses the AEE information 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.[66] When benefit adjustments are made retroactively, it typically means there has been an overpayment of benefits during the prior period of offset use.[67]

As described below, so far less than half of treatment subjects who have used the offset entered through this front door process; the rest entered through the back door. These beneficiaries had above-BYA earnings that they did not report to SSA. Instead, SSA discovered unreported earnings from a different source, most often IRS earnings data (i.e., W-2 reports of earnings). The administrative processes for entry into the offset through the back door differ from the previous list for the front door. Specifically, on the back door path, the third milestone, AEE submission, is not a necessary step. Instead of using an AEE, SSA uses IRS earnings records to retroactively adjust benefits either through an automated process run through BSAS or by a manual process. It is possible for SSA to identify some T1 subjects with unreported earnings before end of year reconciliation makes that person a back door entrant—in which case the back-door entrant may have an opportunity to submit an AEE within the same calendar year as the first month of offset use.

To summarize, in order for SSA to adjust benefits under the offset rules (treatment subjects) or suspend benefits per current law rules (control group subjects), SSA must complete a work CDR based on documented earnings. Such documentation can occur either when the beneficiary reports earnings to SSA or when SSA reviews IRS records. Because it takes considerable time to receive and process the required

[65] Throughout this report we define "offset use" as *eligible and earnings-qualified for the offset, regardless of whether SSA has yet adjusted benefits according to the offset rules*. A beneficiary is considered to have used the offset if the beneficiary has completed the TWP (and thus entered the BPP) and GP *and* has calendar-year earnings that exceed the BYA.

[66] If the first month of the BPP that comes after the GP is any month other than January, then the first possible period of offset use is a partial year (i.e., the remainder of the calendar year) rather than a full calendar year. For earnings to qualify for benefit adjustment in a partial year, the earnings in the remainder of the calendar year must exceed a level equal to the BYA pro-rated by the number of months remaining in the year.

[67] Delays in adjustment may cause underpayments for treatment subjects who were in the EPE and had their benefits suspended before BOND random assignment.

information, SSA often applies the first benefit adjustment retroactively, back to the first month in which the offset use or suspension should apply. SSA may implement these retroactive adjustments many months or even years after the earnings that triggered the adjustment occurred.

7.1.2. Evidence of Progress Toward Benefit Adjustment Under Offset Rules

The previous section described the processes for front door and back door entry into the offset. In this section, we use BTS records to describe the progress of T1 subjects through these milestones in the first five calendar years of the demonstration. Exhibit 7-1 shows the fraction of T1 subjects reaching each of those milestones by December 2015.

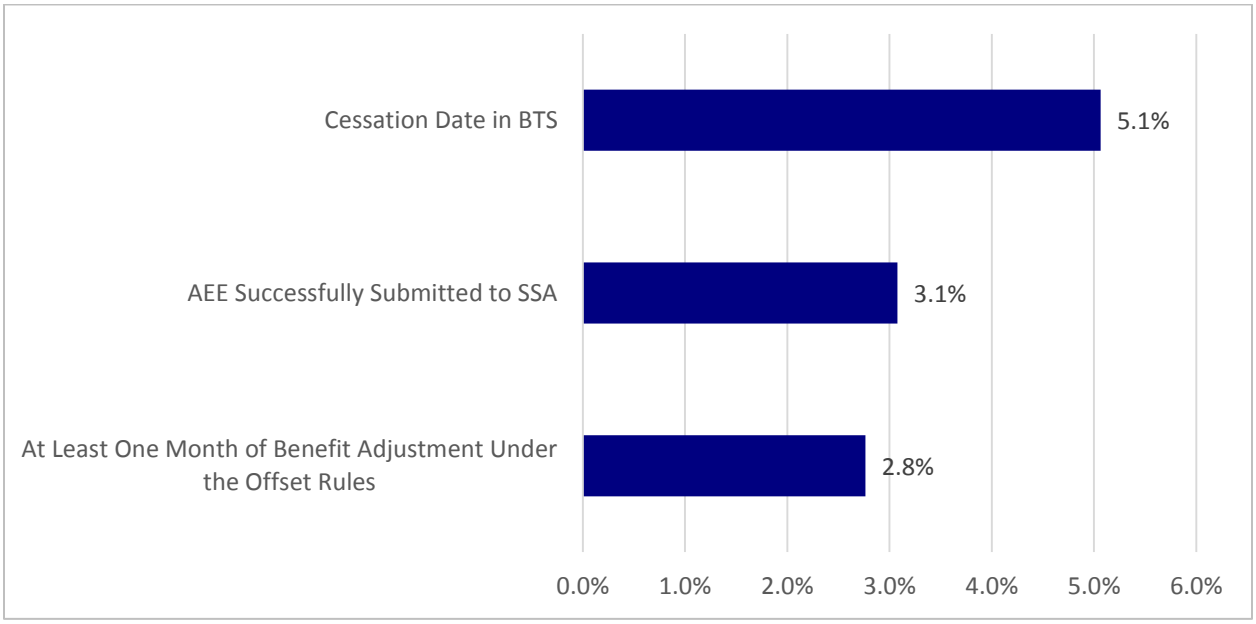
As of December 2015:

- 5.1 percent of T1 subjects had completed the first two milestones—either through the front or back doors. Specifically, they had sufficient earnings to use the offset and SSA had completed a work CDR.
- 3.1 percent of T1 subjects had completed the third milestone; that is they had successfully submitted an AEE to SSA.[68] Everyone in this group had completed the first two milestones (that is, they are included in the 5.1 percent in the first bullet).
- 2.8 percent of T1 subjects had had their benefits adjusted by SSA. Everyone in this group had completed the first two milestones (that is, they are included in the 5.1 percent in the first bullet); but not everyone has completed the third milestone (that is, some are and some are not included in the 3.1 percent in the second bullet).

These proportions for the 2011 to 2015 time period may increase somewhat as SSA completes documentation of T1 work activity for this period.

[68] AEEs are considered successfully submitted once they are acceptable for use by BSAS. In rare cases BSAS does not accept AEEs (for example if an AEE is submitted for a BOND subject who does not yet have a work CDR indicating the beneficiary's disability ceased due to work).

Exhibit 7-1. Percentage of T1 Subjects with Documented Steps toward Benefit Adjustment (through December 2015)



Source: Analysis of BTS records.

Note: According to an analysis of BTS records, 242 beneficiaries successfully submitted an AEE and/or had an adjustment of benefits under the offset rules but did not have a cessation date. 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 individuals constituted 0.3 percentage points of the overall 5.1 percent reported.

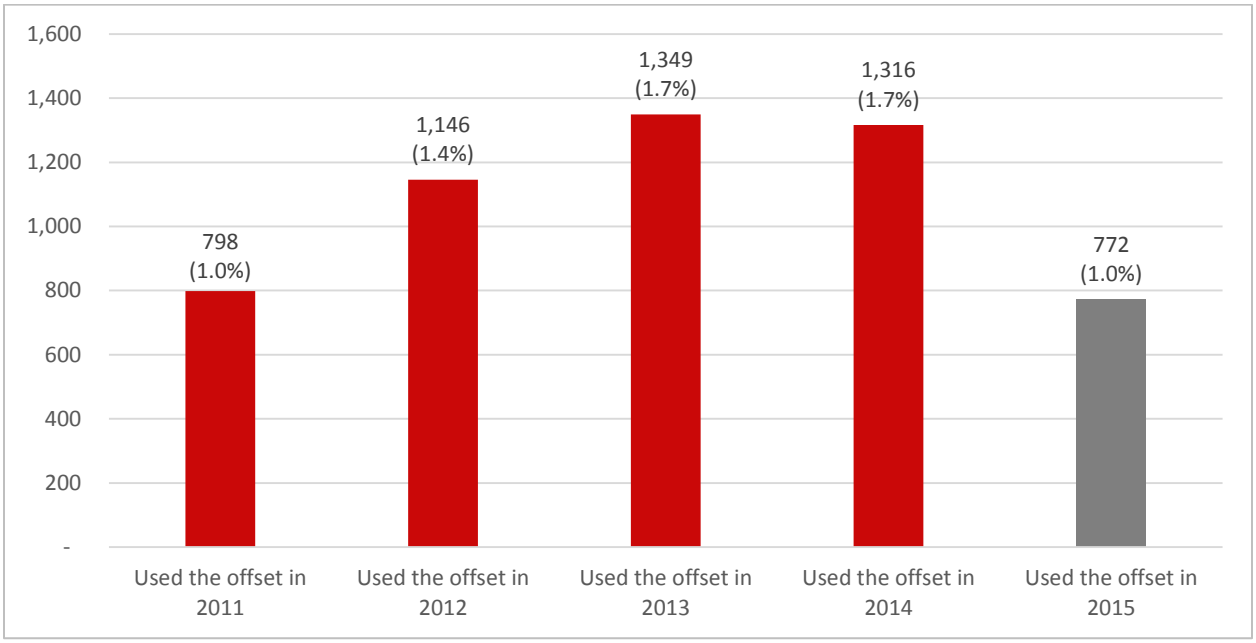
Based on December 2015 data, the number of T1 subjects who had a benefit offset adjustment (the fourth milestone) grew in each of the first three calendar years of the demonstration, peaking at 1,349 T1 subjects—1.7 percent of T1 subjects (Exhibit 7-2). This share was similar in 2014. Based on data through the end of December 2015, as of that same date SSA had applied benefit offset adjustments to the benefits of 1.0 percent of T1 subjects. For all years, we expect that the number of beneficiaries with offset adjustments will increase as SSA retroactively identifies offset users. The count of 2015 offset users—772 T1 subjects according to December 2015 data—will rise the most, largely because SSA has yet to conduct 2015 automated reconciliation (schedule to occur in August 2016). It is possible that SSA will eventually adjust the benefits of more treatment subjects in 2014 or 2015 than in 2013.

As of December 2015, at least 52 percent of the known offset users entered the offset through the back door.[69] This proportion will likely increase for the period through the end of 2015 as SSA continues to retroactively identify T1 subjects with sufficient earnings to use the offset.

[69] Back door offset users are not flagged as such in BTS or SSA administrative data. This 52 percent figure is based on the number of beneficiaries for whom SSA used BSAS without an AEE to make initial adjustments. All of these entrants, plus potentially some whose first adjustment followed a completed AEE, are back door entrants.

As documented in Section 7.5, many entrants experience long delays from the first month in which the offset should be applied to the month in which SSA first adjust benefits. Although delays are typically shorter for those who enter with an AEE—mostly those who enter via the front door—than for those who do not, lags still often are long. For all offset users with first adjustments in 2013 through 2015, the median time from first offset use to first benefit adjustment was 22 months, or just less than two years. These long lags are important for two reasons: (1) they mean that the evaluation does not yet have the full picture of the offset use that has occurred during the demonstration; and (2) they may negatively affect beneficiary understanding of how the offset works.

Exhibit 7-2. Counts of Offset Users by Year, based on December 2015 Data



Source: Analysis of BTS records.

Notes: The automated reconciliation of 2015 earnings has not yet occurred and we show the 2015 counts in gray to distinguish these counts from the previous years' counts that include offset users identified through automated reconciliation. BTS information on years of offset use are missing for 52 T1 offset users.

7.1.3. Organization of the Remainder of the Chapter

The rest of this chapter provides information on the implementation of the BOND demonstration at each of the four milestones along the front door pathway to adjustment of benefits under the offset rules: engagement in sustained earnings, work CDR completion, AEE submission, and first benefit adjustment under the offset rules. The next four sections of the chapter consider operational factors that may have influenced the progress of beneficiaries through these milestones. The following section presents evidence on variations in the pattern of beneficiary progress toward the offset across sites and beneficiary types. A subsequent section addresses the timing and duration of offset use once underway, followed by a concluding section that summarizes the results.

This chapter draws on both quantitative and qualitative data. The quantitative data is primarily from BTS and provides information on the timing of offset use and benefit. When available, we supplement BTS

data with SSA administrative data. The qualitative data were generated in two phases. First, in late 2014 and early 2015 the Evaluation Team interviewed WIC and EWIC supervisors and counselors; a technical assistance provider; the BOND lead for post-entitlement work; the liaison between the BOND sites and the BOND Implementation Team; the Implementation Team director and deputy director; and staff in SSA's ORDES BOND work unit. In late 2015, we collected substantial information about beneficiary experiences from in-depth telephone interviews with 30 work-oriented T1 subjects and follow-up conversations with staff in the ORDES BOND work unit within SSA.

7.2. Activities Related to Maintaining Sufficient Earnings to Use the Offset

The first milestone towards using the offset is sustained earnings. BOND treatment subjects must engage in SGA for a sustained period to use the benefit offset. Specifically, beneficiaries must complete their TWP and GP and then earn more than BYA during a calendar year to qualify for an adjustment of their SSDI benefits for that year according to the offset rules. Factors related to the demonstration as well as a beneficiary's personal circumstances can affect sustained employment at a substantial level and, consequentially, offset use.

One factor influencing offset use may be beneficiary understanding of the work incentives. Implicit in the logic of BOND is that beneficiaries need to understand the benefit offset offer in order to change their behavior in response to the new work incentive. Stage 1 36-Month Survey findings suggest that less than one third (29 percent) of T1 subjects understand that the benefit offset allows them to keep some of their SSDI benefits if they engage in SGA after the TWP and GP (Section 5.1.2). As a result, some T1 subjects may be less likely to take advantage of the opportunity than if they had a better understanding.[70]

The availability of employment services may also influence employment. Some beneficiaries seeking to return to substantial employment need employment-related services to facilitate their efforts. All SSDI beneficiaries, including BOND control subjects are eligible for assistance with work incentives and return to work through SSA's WIPA program. BOND treatment subjects had access to similar services through WIC. Apart from benefits counseling and referrals, employment services are not a component of BOND; treatment subjects presumably have no more or less access to such services than control subjects. BOND work incentives counselors may refer treatment subjects to existing service providers, as WIPA counselors may do for control subjects. However, through 2015, treatment subjects (and presumably control subjects) have not always been able to obtain employment services when needed for reasons external to BOND (Sections 3.5 and 6.2).

Additionally, beneficiaries' personal circumstances may influence employment and earnings. According to the Stage 1 36-Month Survey, nearly 90 percent of T1 subjects cited physical or mental health conditions as barriers to work (Section 6.1). Similarly, during in-depth interviews, seven of the ten work-

[70] As described in the *Process Study Report*, the Implementation Team notified T1 subjects of their involvement in BOND via a letter in 2011. SSA then sent a letter to T1 subjects, informing them about the offset and services available through BOND, including how to contact the demonstration. In 2012 and again in 2013, the Implementation Team conducted follow-up outreach to T1 subjects not yet engaged with the demonstration. At any point, a T1 who contacted the demonstration in response to this outreach or due to awareness of the demonstration by some other means received an explanation of the BOND reporting requirements and the available WIC services.

oriented T1 subjects with a cessation date who had not had a benefit offset adjustment named physical or mental health issues as the main factor determining whether and how much they worked.

7.3. Activities Related to Work CDR Completion

The second step toward adjustment of benefits under offset rules is completion of a work CDR. SSA conducts a work CDR to evaluate a beneficiary's work history and earnings to determine whether or not the beneficiary has completed the TWP and subsequently engaged in SGA. When conducting the work CDR for a BOND treatment subject, SSA identifies the first month in which SSDI benefits should be adjusted according to the offset because of earnings. Benefit adjustment may be delayed, however, because of impediments in identifying beneficiaries who engaged in SGA and completing documentation required to establish offset use via the work CDR process. This subsection describes the work CDR process and considers factors that may have contributed to delays in work CDR completion.

There are three steps in the work CDR process which we discuss in operational terms below: (1) SSA or BOND staff identify those in need of a work CDR based on beneficiary-reported earnings information or other sources, typically an SSA-initiated review of IRS records, (2) beneficiaries, often with the help of SSA or BOND staff, compile information on work history, and (3) SSA verifies the information and completes the work CDR. Several processes in this flow differ across treatment and control subjects in ways that might lead to systematic differences in the timeliness of work CDR completion for the two groups.

As described in this section and in previous reports (Gubits et al. 2013, Derr et al. 2015), these processes involve various SSA and BOND staff. To clarify the discussion, it is helpful to recognize that SSA staff involved in BOND operations are in several different components of SSA. First, the ORDES BOND work unit provides operational support for BOND. This involves collecting work CDR documentation from beneficiaries, assigning work CDR cases to other SSA components for processing, and directly processing the balance of work CDR cases. Second, staff at local SSA field offices are available to assist BOND treatment subjects with activities such as collecting work reports, assisting with work CDR paperwork, and processing select work CDRs. Staff at SSA processing centers also process work CDRs for some treatment subjects. In addition, BOND staff—specifically WIC and EWIC staff and members of the BOND Implementation Team—are available to collect beneficiary work reports and ask the ORDES BOND work unit to initiate a work CDR. As described below the alignment of these responsibilities has evolved over the course of the demonstration.

7.3.1. Identifying Beneficiaries in Need of a Work CDR

The first step in the work CDR process is for SSA to identify beneficiaries who need a work CDR. The BOND Implementation Team notifies SSA about beneficiaries whom it believes require work CDRs. SSA can also identify beneficiaries who require work CDRs using information from IRS earnings data without input from the BOND Implementation Team.

The process for identifying beneficiaries in need of a work CDR differs depending on whether beneficiaries report their earnings. SSA requires all SSDI beneficiaries to report earnings. C1 subjects are required to provide these reports directly to SSA, while BOND treatment subjects may report earnings directly to SSA or they may report earnings to BOND staff. Treatment subjects who report earnings to either BOND or SSA staff are front door entrants to benefit adjustment under the offset rules.

According to members of the BOND Implementation Team and as indicated in the *Stage 2 Early Assessment Report*, because of the complexity of the task and competing demands on their time BOND staff initially struggled to timely identify treatment beneficiaries with reported earnings suggesting need of a work CDR. In response, in March 2013, the BOND Implementation Team began monthly reviews of BTS data to identify treatment beneficiaries in need of a work CDR. In contrast, SSA staff screen work reports as they are submitted.

SSA's guidance to BOND staff about the timing of work CDR initiation changed during the demonstration and at times differed from guidance issued to non-BOND staff. This potentially affected initiation of the work CDR process differently for T1 subjects relative to C1 subjects. Prior to spring 2014, BOND staff were to request a work CDR if a beneficiary was known to have earnings over the SGA amount and was likely to have completed seven or more TWP months. After spring 2014, in an effort to reduce overpayments, which grow as delays in CDR completion lengthen, SSA changed its guidance to BOND staff. The new instructions conformed to the general SSA guidance for current-law beneficiaries: request initiation of a work CDR whenever a beneficiary reports new work at any level. SSA field office staff and processing center staff can initiate this process.

The extent to which the difference in initial guidance affected the timing of work CDRs is unclear. Because C1 subjects could only report earnings directly to SSA, while T1 subjects can report to SSA or BOND staff, the initial work CDR guidance, if followed exactly, was likely to have resulted in delayed work CDR initiation for T1 subjects relative to C1 subjects. However, in late 2015 ORDES staff reported that, in practice, there was little distinction between the standard SSA process for identifying needed work CDRs (applicable to C1 subjects and T1 subjects throughout the demonstration) and the differently designed process operated initially by BOND staff (applicable only to T1 subjects in the first three years of the demonstration). ORDES work unit staff reported that, to promote efficiency, many workers in the SSA process developed a work CDR only when a beneficiary was thought likely to be working at or above the SGA amount, similar to the original BOND process. Even so, the SSA practice of identifying beneficiaries in need of a work CDR is not contingent on beneficiary completion of seven TWP months, so differences may have remained.

Whether subject to current law (SSA 2011) or subject to BOND offset rules (Derr et al. 2015), many beneficiaries fail to report work. For beneficiaries who do not report work, the work CDR process starts when earnings information is received from another source—typically IRS earnings records as indicated on annual tax filings. SSA reviews IRS earnings records three times a year, typically starting in late spring or summer of the following calendar year. Hence, SSA first identifies beneficiaries with unreported earnings about 6 to 18 months after the earnings occurred. Treatment subjects who do not report sustained earnings are back-door entrants to benefit adjustment. Relative to those using the front door (i.e., reporting earnings directly), beneficiaries using the back door face a delay in initiation of the work CDR process.

7.3.2. Developing a Beneficiary's Work History

Once SSA determines that a work CDR is needed, SSA staff send the beneficiary administrative forms and a request to document past work activity. In many cases, beneficiaries need assistance in completing these forms. Both BOND staff and SSA field office staff are available to provide treatment subjects with assistance, if requested. WIPA counselors and SSA field staff may provide similar assistance for C1 subjects.

According to both BOND and ORDES staff, the process of collecting work history information from beneficiaries generally operates well (Derr et al. 2015). However, there are some exceptions. In some cases, beneficiaries may not provide information in a timely fashion (Derr et al. 2015). During 2015 interviews, ORDES and BOND staff shared reports of SSA field office staff declining to provide assistance with paperwork to treatment group beneficiaries because of their BOND treatment subject status. ORDES BOND work unit staff reported that, towards the end of 2014, they had begun to take corrective action to address this issue.

7.3.3. SSA Processing of Work CDRs

Once SSA receives work history information from a beneficiary, SSA staff need to complete three steps: (1) evaluate the earnings documentation; (2) render a decision about whether and when a beneficiary's disability ceased due to work activity; and (3) prepare relevant forms and notices. Since the start of the demonstration, SSA has had a significant backlog completing these steps for treatment subjects and potentially for control group subjects, though the Evaluation Team did not collect qualitative information on lags in the current law process. To some extent, BOND inherited work CDR delays because some BOND subjects were already overdue for work CDR evaluation before the demonstration began. In fiscal year 2010—before BOND began enrolling subjects—SSA took 124 days to process work CDRs on average (SSA 2011). It appears that long processing times persist for treatment subjects for reasons explained below. WIC and EWIC staff cite delays in work CDR processing as a key barrier to timely benefit adjustment.

Following reported backlogs in work CDR processing, starting in 2013, SSA dedicated additional resources to assist the ORDES BOND work unit with processing treatment group work CDRs. As described in the *Process Study Report*, ongoing assistance from both SSA processing centers and field offices has helped to reduce the work CDR backlog. Currently, ORDES processes cases expected to result in disability cessation and sends cases not expected to result in cessation to SSA processing centers or field offices for processing. Under this system, if SSA processing centers and field offices process a work CDR that results in cessation, they transfer it to the ORDES BOND work unit for final processing.

Changes within the ORDES BOND work unit have, however, reduced the internal resources available to process work CDRs. During 2015, SSA reduced the ORDES staff processing work CDRs from six full-time staff to two full-time and one part-time staff.[71] Given these staff cuts and no indication of a change in workload, it is not surprising that in late 2015, BOND work unit staff reported that they had insufficient staff to process BOND work CDRs on a timely basis. Presumably as a result of this under-staffing, the BOND work unit queue had 569 cases in need of full work CDR processing and 252 cases with cessation dates in need of final processing. We do not have comparable estimates of the workload for SSA field offices. However, the work unit staff reported that their workload is much higher than that of their field office counterparts, who are responsible for conducting work CDRs for current-law beneficiaries, including C1 subjects.

[71] The number of staff responsible for processing work CDRs temporarily increased from three staff to five during the summer and then in November 2015 declined to one full-time and one part-time person, both of whom also fulfilled other responsibilities in addition to processing work CDRs.

Aside from insufficient resources to process work CDRs promptly, coordinating SSA field offices and processing centers with BOND benefits counseling staff and BOND treatment subjects for the sake of work CDR processing has been challenging. According to WIC and EWIC staff, the process and the multiple points of contact are confusing to some beneficiaries. Some counselors also expressed frustration that they are unsure where work CDRs are being processed. The same staff noted that they are sometimes unable to ask questions or get status updates from the appropriate source. This problem appears to be most common in cases being processed outside of the ORDES BOND work unit. ORDES staff noted that they are able to provide status updates, including for work CDRs being processed by SSA field offices and processing centers. It is unclear whether some WIC and EWIC staff are unaware of this assistance, find it burdensome to make the request, or find the ORDES updates to be less timely or reliable than desired.

Another challenge is that SSA processing center and field office staff do not have direct access to beneficiary work reports stored in BTS. Two members of the Implementation Team explained that WIC and EWIC counselors must ask ORDES BOND work unit staff to make this information available to the SSA staff processing the work CDR. This adds to the workload of the BOND work unit staff. In addition, if WIC and EWIC counselors are not aware of the ongoing work CDR or otherwise are not proactive about requesting the transfer of BTS information, the staff processing the work CDR at SSA will duplicate previous efforts to collect information (including beneficiary efforts to provide the information) or conduct the work CDR with limited information.

The overall effect of changing resources on work CDR processing times is unclear, but recent statistics suggest that treatment group work CDRs are still subject to long processing times. According to snapshots from SSA's eWork system, between February and December 2015 (approximately the same period as the ORDES staffing reductions reported above), the percentage of BOND treatment group work CDR cases more than 270 days (nine months) old grew from 56 to 71 percent. The corresponding figure for beneficiaries subject to current-law (non-BOND cases and control group subjects) is 2 percent. Thus, it appears that work CDR processing times are much longer for BOND treatment subjects than for control group subjects.

Delays in work CDR completion have implications for benefit adjustment. As a result of delays in activities related to work CDR completion, for the 1,247 cessation dates recorded in BTS between March 2014 and May 2015, SSA assigned only 4 percent of cessation dates for treatment group subjects within three months of cessation. We do not have information on the proportion of C1 beneficiaries who were assigned cessation dates within three months of cessation. Nonetheless, three months is a notable milestone because, including the month of cessation, this is the GP during which benefits are protected from adjustment due to engagement in SGA. The implication is that adjustment of the benefits of almost all treatment group subjects who engaged in SGA after disability cessation has not been timely and thus could result in overpayments as discussed in Chapter 8.

7.4. Activities Related to AEE Submission

The third milestone on the front door pathway to benefit adjustment under offset rules is completion of an AEE. Under current law, after beneficiaries use the three GP months, SSA suspends the benefits of those it determines to be engaging in SGA. In contrast, under the offset rules, after the GP months SSA instead adjusts benefits based on estimated earnings for the remaining months of the calendar year. Completion of an AEE is thus unique to treatment group subjects.

Accurate and timely AEE completion is a necessary step for proper prospective benefit adjustment and help beneficiaries minimize over- and underpayments. When SSA completes a work CDR and identifies a treatment subject earning above BYA after the GP, SSA suspends benefits until the beneficiary submits an AEE.[72] After the first AEE, BOND staff advise beneficiaries to submit an AEE for each calendar year and strive to collect AEEs before the start of a new calendar year. If a beneficiary does not submit a new AEE, SSA will adjust benefits for the coming year in accordance with the most recent AEE submitted.

The processes for identifying beneficiaries in need of an initial AEE and for completing AEEs have evolved over the course of the demonstration. Early in the demonstration, BOND field staff had difficulty identifying beneficiaries in need of an initial AEE. However, since the Implementation Team began monthly reviews of BTS data in early 2013 this process has run smoothly (Derr et al. 2015). Beneficiaries who submit an AEE for the first time do so with the help of a WIC, EWIC, or BOND Implementation Team member, in person or over the phone. For subsequent years, demonstration staff have attempted to collect AEEs by mail.

As discussed in Section 4.2, early in the demonstration, staff struggled to accurately complete AEEs while fulfilling their other post-entitlement work responsibilities. Some staff became proficient with AEEs; for other staff, problems persisted through fall 2013. To address this issue, in December 2013 the Implementation Team centralized post-entitlement work for the majority of the ten BOND sites (Derr et al. 2015). In these sites, members of the BOND Implementation Team, known as the Centralized Post-Entitlement (CPE) Team, complete AEEs.

Centralizing post-entitlement work has led to several improvements, most notably in the quality of AEEs. Members of the BOND Implementation Team explained that post-entitlement work is nuanced, technical, and complex. Many of the CPE team members now responsible for the work have experience working at SSA and, in large part as a result of that experience, have the necessary skills to perform the work. Furthermore, while some WIC and EWIC staff had expressed disinterest in this work, this sentiment appears to be rare among the CPE team. In fall 2014, a large majority (86 percent) of WIC and EWIC focus group participants in centralized sites favored centralization of AEEs and other post-entitlement work. BOND staff perceived that beneficiaries received higher quality support as a result and the staff submitted fewer records with errors. Indeed, in late 2012 and early 2013, about 30 percent of submitted AEEs contained errors (Derr et al. 2015). In 2014, only about 1 percent of initial AEEs contained errors. Although comparable statistics across other years were not available, BOND staff perceived that the timeliness of AEE submission and post-entitlement work had improved, and that the improvement was due, in part, to the streamlined process. ORDES BOND work unit staff agreed that the centralized process was more streamlined and noted that they received fewer trivial questions from staff responsible for post-entitlement work. In addition, work unit staff observed that over- and underpayments resulting from inaccurate AEEs had declined.

[72] AEEs are used for prospective adjustments and thus back-door offset entrants do not have the opportunity to submit an AEE for their first year (or years) of offset use if those years have already passed. These beneficiaries must submit an AEE for the current calendar year and, once notified of this requirement, are subject to benefit suspension until they submit an AEE.

Although centralization of AEEs and other post-entitlement work has yielded mostly positive results, centralization has also created new challenges. First, WIC, EWIC, and CPE team staff indicated that some beneficiaries are confused about the roles of various staff and to whom to direct questions. Indeed, in sites in which the Implementation Team centralized post-entitlement work, beneficiaries may interact with their designated counselor as well as a CPE team member. Implementation Team leadership has instructed WIC and EWIC staff to answer post-entitlement questions and then refer beneficiaries to a CPE team member for further assistance.

Second, in centralized sites, some WIC and EWIC staff reported that they are inadequately informed about the status of the post-entitlement work, and therefore poorly positioned to answer beneficiary questions or anticipate delays or other issues. Each centralized site has a designated CPE team member to handle post-entitlement work, so staff know to whom to direct inquiries, but some counseling staff described being less informed than when they had direct responsibility for the work.

7.5. Activities Related to Benefit Adjustment Under the Offset Rules

Initial benefit adjustment according to the offset rules is the final milestone of the benefit offset adjustment process. SSA developed BSAS, a computer program that interfaces with SSA's data systems, to adjust SSDI benefits for treatment subjects after the beneficiary submits an AEE. SSA also uses BSAS to conduct automated reconciliation with IRS data. BSAS functions as intended for the cases with an AEE but automated reconciliation functionality has been problematic.

7.5.1. Implementation of The Benefit Adjustment Process

The process used to adjust benefits under offset rules differs according to whether or not the beneficiary submits an AEE. Front-door offset users submit AEEs ahead of their initial offset adjustment. Depending on the timing of work CDR completion relative to the date the beneficiary's earnings first meet the threshold for offset use, these adjustments may take place during the initial calendar year of offset use or during a subsequent calendar year.[73] In addition, both front door- and back-door offset users generally submit AEEs to facilitate adjustment for years of offset use following their initial adjustment. In all of these cases for which the AEE predicts earnings above BYA, BSAS uses the information on the AEE to make contemporaneous adjustments to benefits. In addition, BSAS makes adjustments in response to revised AEEs submitted within the calendar year.

After the end of each calendar year, SSA also uses BSAS to compare expected earnings to earnings reported in IRS records and makes additional retroactive benefit adjustments for the prior year in the event of a substantial difference. The purpose of these reconciliations is to issue the correct benefit amount for the previous calendar year. SSA only processes automated reconciliation for beneficiaries with a work CDR determination indicating disability cessation because only those beneficiaries are potentially eligible for adjustment of benefits under the offset. This group includes both back-door offset users who have not previously had a benefit offset adjustment as well as beneficiaries whose benefits

[73] Note that the initial benefit adjustment may not be for the first year of offset use, but instead for a later year of offset use. When this is the case, a reconciliation is required to determine the benefit adjustment amount for each earlier year of offset use.

were adjusted for the year in consideration. SSA uses manual reconciliation to adjust benefits in earlier years for which automated reconciliation has already taken place.

7.5.2. Performance of BSAS

BSAS performs well for most contemporaneous adjustments based on AEEs, with one exception. At the start of the demonstration, AEE-based offset adjustments took no longer than three days. A BSAS correction in December 2012 successfully addressed the underlying issues causing those relatively minor delays (Gubits et al. 2016). According to ORDES and BOND staff, BSAS currently functions well for many adjustments based on AEEs. The exception is AEEs for former offset users whose earnings have dropped below BYA. BSAS is unable to process these cases, and instead ORDES staff must process them manually. The Implementation Team has observed delays with this manual process, leading to delays in beneficiaries' return to full benefits.

For retroactive adjustments made through automated reconciliation, BSAS did not work well during the first three years of the demonstration. One issue was related to the timeliness of automated reconciliation. We have previously reported that issues with BSAS functionality were responsible for long delays in the automated reconciliation processes for 2011 and 2012 benefits, and these issues persisted for the 2013 benefit adjustment process (Derr et al. 2015). Automated reconciliation for a given year is scheduled for August of the following year, but SSA delayed the 2011 automated reconciliation by five or six months (conducted in January and February 2013) and 2012 automated reconciliation by one or two months (conducted in four batches in September and October 2013). SSA conducted 2013 automated reconciliation in late April through May 2015—an eight month delay. The direct result of such delays is an extended wait for benefit reconciliation, which affects both beneficiaries with a benefit offset adjustment in the previous year and first-time offset users. For the latter group, the result is an extended period of time over which beneficiaries may accumulate over- and underpayments and perhaps not understand how their earnings are affecting their benefits. These delays have presumably been more common for back door entrants than for front door entrants, because back door entrants typically enter through the automated reconciliation process that was the source of many of the adjustment delays.

Going forward, the timeliness of automated reconciliation appears to have been resolved. SSA ran automated reconciliation for 2014 on schedule in August and September 2015 and expects to adhere to the same schedule for automated reconciliation for 2015 earnings.

In addition to the delays in initiating automated reconciliation, BSAS was not fully automated for many cases. During the 2014 automated reconciliation, BSAS could not process more than 3,000 cases. This represents more than half of all cases included in the 2014 automated reconciliation.[74] In most cases, the consequence of the error was that SSA sent an overpayment notice a month or two after the automated reconciliation. ORDES work unit staff identified at least one underlying cause of the processing limitations, and expect a correction of the problem to reduce manual corrections in future automated reconciliations by a third. As with delays in work CDR processing at the work unit, staff ability to manually process cases that BSAS cannot process and to diagnose the problems in BSAS is limited by

[74] According to BTS data from late August 2015 (around the time of 2014 automated reconciliation), 5,098 treatment subjects had cessation dates and thus were presumably included in this process. This includes both Stage 1 and Stage 2 treatment subjects.

insufficient resources relative to their workload. Furthermore, this detracts from time that could otherwise be spent processing work CDRs.

Although most BSAS limitations affect only the timing of the completion of automated reconciliation, one limitation was known to produce incorrect calculations of underpayments for beneficiaries with consecutive years of underpayments. This issue affected approximately 200 subjects in 2011 and 300 subjects in 2012 and may have led to erroneous notices of underpayment for those cases, according to ORDES staff. These cases also require manual review. The issue is ongoing, although ORDES staff now review these cases to ensure that they send correct notices to affected beneficiaries.

7.5.3. The Timing of Benefit Adjustment

Previous sections examined the factors influencing attainment of offset milestones. Here we look at when first benefit adjustments actually occurred and examine the duration from first month of offset use to the month in which SSA first makes an adjustment in benefits. These statistics reflect the aggregate effects of factors affecting the speed of the adjustment process.

Exhibit 7-3 compares cumulative statistics on first month of offset use to cumulative statistics on the first month SSA adjusted benefits, based on SSA administrative data and BTS records.[75] The upper line in the exhibit shows the cumulative percentage of T1 subjects who were known to have first begun a period of offset use (i.e., earned above BYA during the BPP) in the indicated month, based on adjustments completed through the end of 2015. This line will change as SSA makes more retroactive adjustments in the future.

The lower line in the exhibit shows the cumulative percentage of T1 subjects for whom SSA had actually made an initial adjustment by the indicated month. This series will not change as SSA makes retroactive adjustments. In a few months this series declines, because SSA reversed some adjustments after the initial adjustment is made.[76] The main reason the two series differ is that SSA's initial adjustment of benefits for a beneficiary generally occurs some amount of time after the first month of offset use.[77]

An example is helpful in distinguishing between the two cumulative series. If a beneficiary had a cessation date and sufficient earnings to first use the offset in May 2012, but SSA adjusted his or her benefits in September 2013 (retroactively), the beneficiary would be included in the upper line starting in

[75] The data on cumulative offset use are based on a monthly extract from SSA's MBR as well as calculation and verification of first offset month by SSA staff. The data on cumulative percentage with adjustments in 2011 and 2012 are from manual updates made by SSA staff to BTS and were verified by SSA staff. The data on cumulative adjustments in 2013, 2014, and 2015 are from a combination of BTS, BSAS, and MBR data.

[76] Such changes affect the cumulative offset use series (the upper line) in a different way—the whole series is reduced starting with what was incorrectly considered to be the first month of offset use, so these changes may not be as obvious.

[77] The value of the both series are equal in December 2015 because they are from data in which all those known to have used the offset by December 2015 had their benefits adjusted by December 2015. SSA continues to retroactively identify additional T1 offset users in 2015 or earlier, so the cumulative percentage of T1 subjects who used the offset during the period (including in December 2015) will increase, but the percentage of initial adjustments during the period will not change, by construction.

May 2012 and in the lower line starting in September 2013. In contrast, if the same beneficiary had entered through the front door and SSA had first adjusted his or her benefits under the offset rules contemporaneously with the first month of offset use (May 2012), the beneficiary would be included in both lines starting in that month.

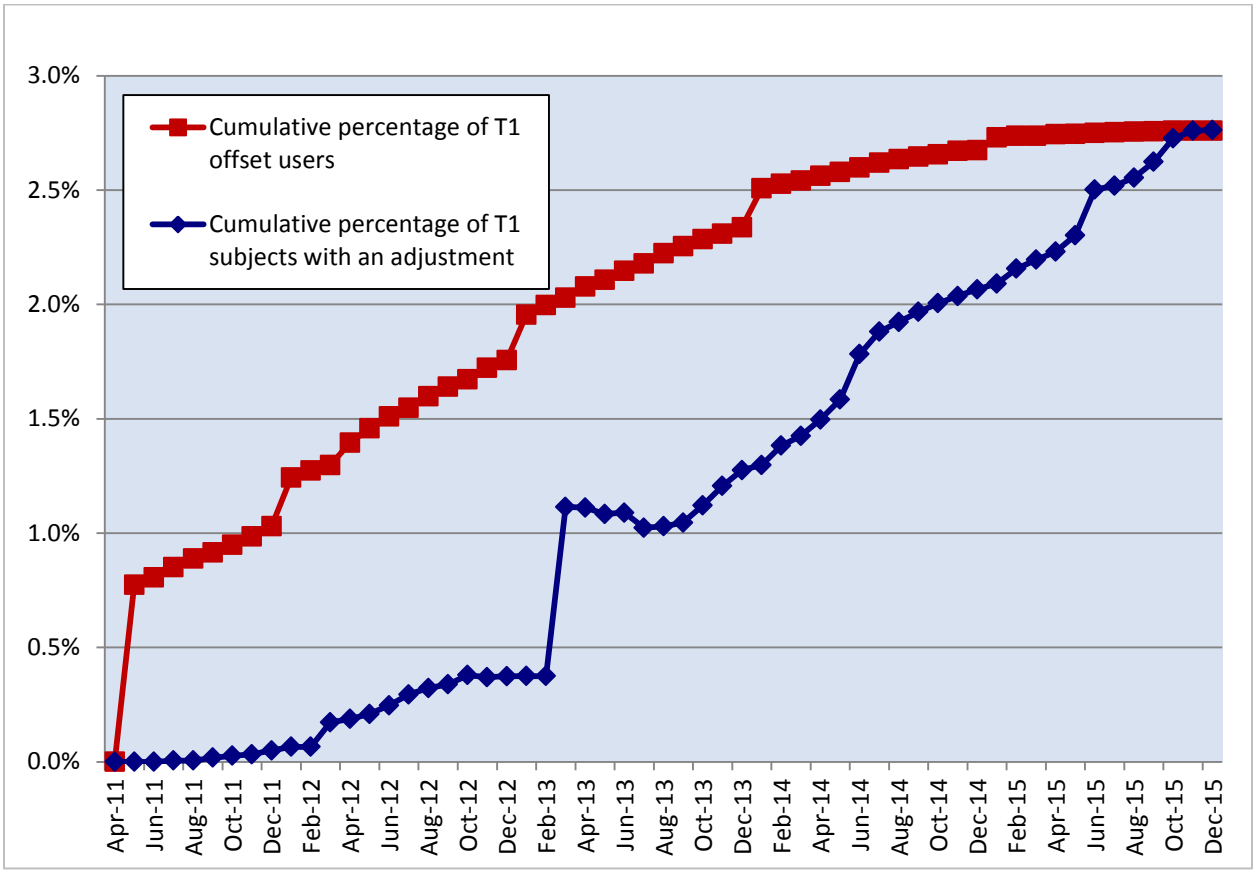
The number of T1 subjects with a benefit adjustment grew throughout the demonstration period, but with a different trajectory compared to the number of offset users. A comparison of the lower line in Exhibit 7-3 to the corresponding upper line shows that SSA was delayed in making the first adjustments for many T1 offset users. For example, in February 2013, 1,586 beneficiaries (2.0 percent) had used the offset but SSA had only adjusted the benefits of 298 beneficiaries (0.4 percent). In the next month, March 2013, the number of offset users grew marginally to 1,612 beneficiaries (2.0 percent) while the number of beneficiaries with an adjustment jumped to 885 (1.1 percent). The jump in benefit adjustments in March 2013 is a result of SSA's completion of 2011 automated reconciliation.

Ideally, SSA would first adjust benefits in the first month of offset use, or shortly thereafter. Such rapid adjustments would help beneficiaries understand how their earnings affect their benefits and total income, and minimize variation in the beneficiaries' monthly income due to delays in administrative processes. It is apparent from the above statistics, however, that the duration from first month of offset use to first adjustment is substantial. This duration reflects all delays leading up to benefit adjustment including beneficiaries who do not report earnings in the required timely manner and SSA delays in processing work CDRs and automated reconciliations.

Here we provide statistics on the duration between the first month of offset use and SSA's first adjustment of benefits for beneficiaries whose benefits were first adjusted between February 25, 2013 and December 31, 2015.[78] We do not have comparable BTS statistics for C1 subjects (duration from first month of SGA-level earnings after the GP during the EPE until benefits are actually suspended). It is important to keep in mind that the unobserved durations for C1 subjects may be typically as long as or longer than for T1 subjects, because many of the administrative issues that are the source of major delays exist outside of BOND.

[78] Data were available for 87 percent of the 2,195 beneficiaries for whom SSA adjusted benefits by the end of December 2015. SSA adjusted benefits for the remaining 13 percent before February 25, 2013.

Exhibit 7-3. Cumulative T1 Offset Users and Cumulative T1 Subjects With an Offset Adjustment, based on December 2015 Data



Source: Monthly extracts from SSA's MBR.

Note: The upper line, cumulative percentage of offset users, shows the cumulative percentage of treatment subjects who completed the TWP and GP and then earned above BYA in at least one month, based on administrative records through December 2015. The cumulative percentage of offset users at any point in time presented in this series will continue to increase as SSA completes retroactive adjustments for this period.

The lower line in the exhibit, cumulative percentage with an adjustment, provides information on the months in which initial benefit adjustments under the offset rules were made—usually later than the first month of offset use. It represents the percentage of beneficiaries whose benefits actually have been adjusted under the offset rules as of the month indicated. Declines that sometimes occur in the cumulative percentage with an adjustment from one month to the next are due to retroactive reversals of initial adjustments. Such cases are not included in the cumulative percentage of offset users because the action determined that they had not actually used the offset.

For all offset users with first adjustments in 2013 through 2015, the median time from first offset use to first benefit adjustment was 22 months, or just less than two years. To better understand the reason for this substantial delay, we examined duration separately for the two primary processing pathways that treatment subjects may take to first adjustment: (1) submission of a qualifying AEE, a proxy for front-door offset entry and (2) SSA-initiated reconciliation, which may take place through SSA's annual automated review of previous year IRS data for subjects with cessation dates, or manually, a proxy for

back-door offset entry.[79] By design, the first pathway should take less time than the second. This is because AEE submission has the potential to provide more timely earnings information to assess eligibility for adjustment under the offset rules. In addition, beneficiaries whose benefits are adjusted based on the submission of an AEE may be more proactively engaged in the demonstration than beneficiaries whose offset use was discovered via a retroactive review of data. Finally, some adjustments for those who entered via the backdoor were delayed due to previously noted delays in the automated reconciliation process. As a result we present processing times separately by pathway.

As expected, median time from first offset use to adjustment was not as long for beneficiaries whose initial adjustment was in response to a submitted AEE (i.e., front door) as for beneficiaries who entered the offset via reconciliation (i.e., back door). The median processing time was 17 months for adjustments in response to AEEs, relative to 23 months for adjustments resulting from reconciliations. As indicated earlier, 48 percent of adjustments were based on AEEs and 52 percent were based on reconciliations.

The histograms in Exhibit 7-4 show that duration from initial offset use to first adjustment varies substantially around the medians, and point to differences between the two entry paths. SSA adjusted the benefits of 36 percent of those who entered via an AEE within 12 months of the first month of offset use, whereas the comparable figure for initial adjustments via reconciliation is only four percent. At the opposite end of the distribution, the duration for 26 percent of entrants via an AEE was more than 24 months, compared to 36 percent for initial adjustments made via reconciliation. These cases include four percent of entrants via an AEE and eight percent of entrants via reconciliation with durations longer than 36 months.

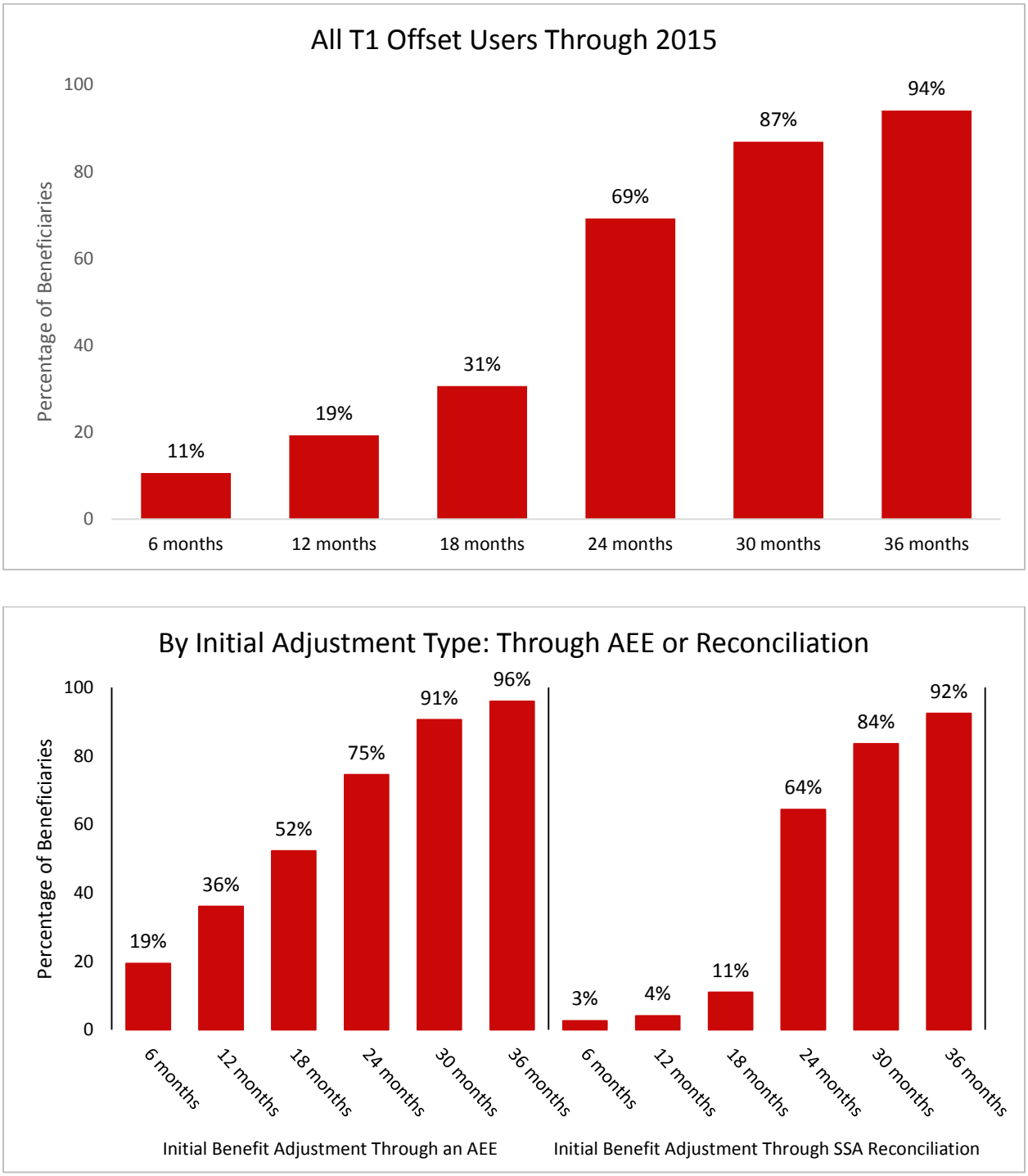
The long durations experienced by many benefit offset entrants are particularly problematic for the demonstration because the period of offset eligibility—the BPP—is limited to the 60 months after TWP completion, or the 60 months after the start of BOND for those who completed their TWP earlier. Hence, if a beneficiary completed the TWP after the start of BOND, immediately used his or her three GP months, first used the offset in month four after TWP completion, but did not experience a benefit adjustment until 22 months later (month 25 of the BPP), the beneficiary would not fully experience the income consequences of his earnings until almost half-way through the BPP. The BOND logic model posits that beneficiaries need to understand the benefit offset in order to change their behavior in response to the incentive. Delays in delivery of the incentive via the adjustment process are noteworthy because they may weaken beneficiary understanding of how the offset works. We do not know the extent to which behavior might have differed had this duration been shorter. These long durations are particularly problematic for back door entrants because it is possible that these subjects had little or no awareness about how the offset would affect their benefits before the adjustment was made.

The above discussion ignores the fact that delays may lead to accumulation of overpayments, and the overpayments themselves may have impacts on subsequent earnings that are difficult to predict,

[79] Manual reconciliation is conducted after the end of the calendar year and may occur either before or after the automated reconciliation. Beneficiaries may request that a manual reconciliation take place before the scheduled automated reconciliation. SSA also conducts manual reconciliations for calendar years in which SSA already completed automated reconciliation. See Section 5.2.2 of the *Process Study Report* for more details.

especially if unexpected. Evidence on the prevalence and size of overpayments for both treatment and control subjects appears in the next chapter.

Exhibit 7-4. Cumulative Distribution of Duration from First Offset Use to First Benefit Adjustment



Source: BTS

7.6. Variation in Steps to Offset Adjustment Across Sites and Beneficiaries

In this chapter, we have described the steps and processes associated with the pathway to benefit adjustment in aggregate. Here, we describe variation in these milestones by site and beneficiary characteristics.

7.6.1. Variation in Steps to Offset Adjustment in Sites

The percentage of T1 subjects for whom SSA had adjusted benefits under the offset as of December 2015 varies from a low of 1.7 percent in Alabama to a high of 5.0 percent in DC Metro (Exhibit 7-5). This variation is also present in shares with cessation dates in the first column of the exhibit and shares with successful AEEs in the second column. With only one exception, the rank of each site by percentage with an offset adjustment is identical to the rank by percentage with a cessation date in BTS, indicating that cessation is an essential precursor to subsequent events —AEE submissions, offset adjustments—in the different sites. Indeed, the number of offset adjustments per case with a cessation date is quite similar across sites (between 0.47 and 0.64).

We have no direct evidence on what causes variation across sites in the percentage of T1 subjects with cessation dates. Possible explanations include factors external to BOND implementation such as local economic conditions (Section 3.3), the availability of adequate employment support services (Sections 3.8 and 6.2), and site-specific differences in beneficiaries' characteristics or preferences, including their timeliness of reporting or readiness to use the benefit offset. Site-to-site differences could also be caused by variations in implementation such as how quickly WIC counselors in each site identify T1 subjects in need of work CDRs.

Exhibit 7-5. Percent of T1 Subjects in Offset by Site, based on December 2015 Data

	Cessation Date in BTS (%)	AEE Successfully Submitted to SSA (%)	At Least One Month of Benefit Adjustment under the Offset Rules (%)	Offset Adjustments per Cessation Date
Alabama	3.4	1.9	1.7	0.50
Arizona/SE California	5.5	3.2	2.6	0.47
Colorado/Wyoming	5.5	3.2	3.1	0.56
DC Metro	7.8	5.3	5.0	0.64
Greater Detroit	4.4	2.6	2.3	0.52
Greater Houston	7.1	4.4	4.1	0.58
Northern New England	6.0	3.8	3.5	0.58
South Florida	4.4	2.3	2.2	0.50
Western New York	4.7	3.1	2.9	0.62
Wisconsin	4.6	3.1	2.4	0.52
Total	**5.1**	**3.1**	**2.8**	**0.55**

Source: Analysis of BTS records.

7.6.2. The Relationship Between Beneficiary Characteristics and Steps Towards Benefit Offset Adjustment

Progress along the pathway to benefit offset adjustment varies with beneficiary baseline characteristics. Exhibit 7-6 compares baseline characteristics of three groups of T1 beneficiaries based on achievement of the different offset milestones through December 2015. The groups are: (1) beneficiaries who had neither a disability cessation date nor a benefit offset adjustment ("non-users"), (2) those with a cessation date but no benefit offset adjustment, and (3) those who had a benefit adjustment in at least one month through December 2015. Compared to both non-user groups combined, at the time of enrollment, beneficiaries with an offset adjustment were more likely than the two non-user groups to be young (20-29 or 30-39 years old at baseline), to have a primary impairment of neoplasms (for example, cancer), and to be the direct recipient of benefit payments (rather than receive benefits through a representative payee). These findings are consistent with a comparison of Stage 2 beneficiaries who had benefits adjusted through April 2013 to non-offset users presented in the *Stage 2 Early Assessment Report*.

Beneficiaries with cessation dates who had and had not used the offset were more similar to each other than to those without a cessation date in at least one regard: concurrent receipt of SSDI and SSI benefits. Between 87 and 88 percent of non-users with a cessation date and beneficiaries with an offset adjustment received only SSDI benefits compared to 83 percent of non-users without a cessation date.

Surprisingly, with regard to several other characteristics, beneficiaries with an offset adjustment were more similar to non-users without a cessation date than to non-users with cessation dates. First, about half of non-users without a cessation date and 54 percent of those with an offset adjustment were short-duration SSDI beneficiaries, while roughly a quarter of non-users with a cessation date were short-duration beneficiaries. Second, non-users with a cessation date were nearly 6 to 11 percentage points more likely to have a primary impairment of mental disorders than beneficiaries in each of the other two groups, which were again more similar in this respect. Finally, non-users with a cessation date had the lowest AIME and monthly SSDI benefit amounts of the three groups.

These results are surprising because we would expect non-users with cessation dates to be more similar to users than non-users without cessation dates. However, the results are consistent with findings for Stage 2 subjects described in the *Stage 2 Interim Report*. Future reports will examine the share of current non-users with cessation dates who eventually have their benefits adjusted and consider whether current non-users with cessation dates are more likely to later have benefit adjustments than non-users without cessation dates.

Exhibit 7-6. Treatment Subject Characteristics by Steps toward Benefit Offset Adjustment (through December 2015)

Baseline Characteristic	Non-offset User, No Cessation Date (1)	Non-offset User, with Cessation Date (2)	Benefit Offset Adjustment by December 2015 (3)	P-value
Number of Beneficiaries	***75,412***	***1,829***	***2,195***	
Gender				
Male (%)	51.6	51.0	51.4	0.865
Female (%)	48.4	49.0	48.6	
Age				
20–29 years (%)	7.3	9.3	18.5	0.000***
30–39 years (%)	12.7	21.5	23.0	
40–44 years (%)	10.4	14.8	13.7	
45–49 years (%)	16.3	17.9	14.5	
50–54 years (%)	23.3	19.1	16.3	
Over age 55 (%)	30.0	17.3	14.0	
Mean age (years)	47.6	44.3	41.7	0.000***
Primary Impairment				
Neoplasms (%)	3.2	2.7	6.4	0.000***
Mental Disorders (%)	29.6	41.2	34.9	
Back or Other Musculoskeletal (%)	24.8	16.9	18.5	
Nervous System Disorders (%)	7.2	5.9	5.3	
Circulatory System Disorders (%)	6.7	3.3	4.7	
Genitourinary System Disorders (%)	1.8	2.2	3.8	
Injuries (%)	4.2	5.1	5.6	
Respiratory (%)	2.2	1.5	1.5	
Severe Visual Impairments (%)	1.9	1.5	2.1	
Digestive system (%)	1.8	1.5	1.7	
Other impairments (%)	16.6	18.2	15.5	
Length of SSDI Receipt				
Short duration (36 months or less) (%)	50.4	23.1	54.6	0.000***
Number of Years Received SSDI	6.4	9.4	4.7	0.000***
Benefit Amount and Status				
Monthly SSDI Benefits ($)	$1,010	$992	$1,058	0.000***
AIME (May 2011) ($)	$1,706	$1,511	$1,922	0.000***
Disabled adult child (DAC) (%)	11.2	5.9	5.1	0.000***
Disabled widow beneficiary (DWB) (%)	1.9	0.7	0.5	0.000***
Dually-entitled disabled adult child (%)	1.9	2.3	2.3	0.698
Dually-entitled disabled widow beneficiary (%)	0.9	0.6	0.4	0.002***
Payee is other than self (%)	17.0	13.8	10.6	0.000***
SSDI-only	82.7	88.1	86.7	0.000***
Concurrent	17.3	11.9	13.3	

Source: Analysis of BTS records and baseline SSA administrative records.

Note: p-values shown are from statistical tests of differences in percentages across the three groups. Groups of mutually-exclusive characteristics were tested for differences with chi-squared tests. Single characteristics not part of a mutually-exclusive group were tested for differences by F-tests. The Chi-squared statistic from the omnibus statistical test of difference between groups across all characteristics is 3,041.35, with a p-value of 0.000.

***/**/* indicate statistical significance at the 0.01/0.05/0.10 levels.

Exhibit 7-7. Predictors of Benefit Adjustment under the Offset Rules through December 2015

Predictor of Benefit Offset Adjustment	Coefficient (1)	Standard Error (2)	P-value
Gender			
Male	-0.29	0.11	0.026**
Age			
20–29 years	7.92	0.69	0.000***
30–39 years	4.62	0.49	0.000***
40–44 years	2.90	0.32	0.000***
45–49 years	1.56	0.22	0.000***
50–54 years	0.83	0.20	0.003***
Over age 55	0.00	--	--
Primary Impairment			
Neoplasms	1.83	0.78	0.043**
Mental Disorders	-0.44	0.24	0.099*
Back or Other Musculoskeletal	-1.01	0.22	0.001***
Nervous System Disorders	-1.61	0.28	0.000***
Circulatory System Disorders	-0.96	0.36	0.025**
Genitourinary System Disorders	1.74	0.56	0.012**
Injuries	0.07	0.29	0.828
Respiratory	-0.95	0.41	0.046**
Severe Visual Impairments	-0.52	0.42	0.254
Digestive system	-0.66	0.41	0.148
Other impairments	0.00	--	--
Length of SSDI Receipt			
Short duration (36 months or less)	-0.40	0.18	0.055*
Number of Years Received SSDI	0.01	0.01	0.313
Benefit Amount and Status			
Monthly SSDI Benefits ($1,000)	0.30	0.24	0.243
AIME (May 2011) ($1,000)	0.38	0.12	0.009***
Disabled adult child (DAC)	-3.13	0.43	0.000***
Disabled widow beneficiary (DWB)	-0.47	0.33	0.188
Payee is other than self	-1.53	0.25	0.000***
SSDI-only	0.92	0.32	0.017**

Source: Analysis of BTS records and baseline administrative SSA records on treatment group subjects (T1).

Notes: Findings were derived from a clustered linear regression model without weights, with a dependent variable indicating whether the beneficiary received a benefit payment under the offset in at least one month through December 2015. Adjusted R-Squared: 0.0189. Model F-statistic is 26.71, p-value 0.000.
Sample size: 79,436.

***/**/* indicate statistical significance at the 0.01/0.05/0.10 levels.

In a multivariate regression analysis, many of these beneficiary characteristics are predictive of being an offset user by December 2015 (Exhibit 7-7). Age is a statistically significantly predictor of benefit adjustment under the offset. For example, beneficiaries ages 20-29 were 7.9 percentage points more likely to use the offset relative to beneficiaries ages 55 and older, holding other characteristics constant. Similarly, primary impairments of neoplasms, and genitourinary system disorders were all associated

with a higher likelihood of offset use relative to beneficiaries with impairments in the "other" category, while mental disorders, back or musculoskeletal disorders, nervous system disorders, circulatory system disorders, and respiratory disorders were significantly associated with a lower likelihood of offset use relative to beneficiaries with "other" impairments. Beneficiaries on SSI for 36 months or less, disabled adult child beneficiaries, beneficiaries with representative payees, and concurrent SSDI and SSI recipients were also less likely to have a benefit adjustment than beneficiaries without those characteristics, all other things equal.

7.7. Timing and Duration of Offset Use

After first using the benefit offset (that is, completing the TWP and GP and earning above BYA during the BPP), beneficiaries may continue to earn above BYA and continue to use the offset or may return to full benefits if earnings are reduced. The latter may occur because of a medical problem or a change in the beneficiary's circumstances, but beneficiaries may also choose to adjust their earnings behavior toward any available earnings/benefit scenario that they prefer. This section presents annual statistics on offset users and on the duration of offset use as well as differences in characteristics of beneficiaries who used the offset for different lengths of time.

Temporal patterns of offset use vary. Of T1 subjects known as of December 2015 to have used the offset at some point between 2011 and 2015, 28 percent used the offset in at least one calendar year (Exhibit 7-8). Progressively smaller percentages used the offset for at least two, three, four, or five years.[80] All T1 subjects included in Exhibit 7-7 may extend their duration of offset use in future years. Additional T1 subjects may also use the offset for the first time in 2016 and beyond, and their behavior will affect future updates to this table.

Exhibit 7-8. Duration of Offset Use, based on December 2015 Data

	Offset User Between 2011 and 2015	
	Count of Beneficiaries	Percentage of beneficiaries
Cumulative number of offset users between 2011 and 2015	2,195	100.0
Offset use in one calendar year	621	28.3
Offset use in two calendar years	529	24.1
Offset use in three calendar years	448	20.4
Offset use in four calendar years	367	16.7
Offset use in five calendar years	178	8.1

Source: Analysis of BTS records.

Note: The percentage of offset users in a calendar year is the proportion of all treatment subjects who earned above BYA after the GP and during their BPP in that calendar year, based on administrative records through December 2015. The cumulative number of offset users between in 2011 and 2015 will continue to increase as SSA completes retroactive adjustments for this period. We do not have data to calculate the duration of offset use for 52 of the 2,195 T1 subjects.

[80] We are unable to calculate the duration of offset use for 52 of the 2,195 T1 offset users known by December 2015 because of missing or incomplete BTS data.

Over the period observed so far, the majority of beneficiaries used the offset continuously rather than intermittently. Indeed, 83 percent (437/529) of beneficiaries who used the offset for two years, 81 percent (364/448) of beneficiaries who used the offset for three years, and 88 percent of beneficiaries who used the offset in four years (323/367) were continuous offset users.

The in-depth telephone interviews conducted in late 2015 with 20 T1 subjects who used the offset provide some insights into why some beneficiaries use the offset only for a short time, while others use the offset for long periods (Section 2.1.1). Among these 20 interviewees, half used the offset for a single year ("short-term offset users") and half used the offset for three or more consecutive years ("long-term offset users"). As noted in Section 2.2.1, findings from these interviews should be interpreted with caution due to the small size of the sample and the fact that the respondents cannot be considered representative of all T1 subjects with cessation dates or of their three subgroups.

Long-term offset users reported different barriers to work and reasons for work than short-term offset users. Respondents with short-term offset use most frequently named physical or mental health issues as the main factor determining whether and how much they worked and earned (7 of 10). Only 2 of 10 respondents with long-term offset use gave the same answer. Instead, long-term offset users were more likely to identify facilitators to work, such as work accommodations, or personal motivation to work and earn more as the main factor influencing whether and how much they worked and earned. However, we cannot determine to what extent these differences: (1) caused the differences in their duration of offset adjustment; or 2) are the consequence of other differences that also affect their eligibility for adjustment under the offset rules.

In response to a separate question, a substantial minority of both groups reported that the possibility of a benefit adjustment (reduction) was also an important factor in their decisions whether and how much to work, describing it as a reason to limit their work or earnings. Among beneficiaries with short-term offset adjustments, 3 of 10 provided such a response, as did 2 of 10 beneficiaries with long-term offset adjustments.

In addition, long-term offset users learned about their offset adjustments from a different source than short-term offset users. The former group was more likely to describe first learning about their benefit adjustment from their BOND counselor (7 of 10) in contrast to the latter group, who were more likely to describe first learning of their benefit adjustment through a letter from SSA (8 of 10). This suggests that short-term offset users are more likely to be back-door offset users, who we presume would learn about their offset adjustment from SSA. In contrast, long-term offset users are more predominately front-door offset users, who consult with their WIC counselors about employment and benefits.

7.8. Summary

As of December 2015, SSA had adjusted the benefits of 2.8 percent of T1 subjects. Another two percent had reached another major milestone on the pathway to the benefit adjustment; that is, SSA had established an SGA cessation date or the beneficiary had successfully submitted an AEE to SSA. The percentage of identified T1 subjects known to have used the offset by the end of 2015 will increase as SSA completes the processing of the backlog of work CDR cases.

The number of offset users recognized to date is a function of several factors. First, some beneficiaries may be unable, uninterested, or unprepared to engage in sustained SGA-level work. Second, those

interested, but unprepared, may require time to obtain counseling, employment-related services, or address a variety of issues, and find an SGA-level job. Third, those who, in response to BOND, promptly initiated SGA-level work needed time to work sufficiently to use the offset. Fourth, once beneficiaries have worked enough to warrant a benefit adjustment under the offset, there is often a delay before SSA actually makes the adjustment. We expect the number of known offset users to continue to rise as SSA completes the benefit adjustment process for subjects with sufficient sustained earnings.

The delays from the month in which treatment beneficiaries first earn enough to warrant an adjustment under the offset have been quite substantial. For all offset users with first adjustments in 2013 through 2015, the median time from first offset use to first benefit adjustment was 22 months, or just less than two years. These long processing times present challenges for both the implementation and evaluation of BOND. In terms of implementation, as discussed in Section 5.1, the BOND logic model posits that beneficiaries need to understand the benefit offset in order to change their behavior in response to the incentive. Delays in delivery of the incentive via the adjustment process may negatively affect beneficiary understanding of how the offset works. In terms of evaluation, long processing times before initial benefit adjustment mean that we do not yet have the full picture of the offset use that has occurred in the 2011-2015 period, particularly in the most recent years. Benefit adjustment delays may also lead to overpayments, which may influence beneficiary behavior and contribute to confusion about the relationship between benefits and earnings. We return to this topic in the next chapter.

We have identified three main sources of delays in the adjustment of benefits. First, the failure of many beneficiaries to report earnings delays the start of the benefit adjustment process (that is, back-door offset adjustment). Second, lags in the processing of work CDRs, once SSA recognizes the need for a work CDR, delay the determination of when the beneficiary first used the offset and provides a longer time frame over which the beneficiary may accrue improper payments. These lags are primarily the result of insufficient resources at the BOND work unit to process the work CDR cases timely. Many BOND treatment subjects in the ORDES work CDR queue (71 percent) encountered CDR processing times longer than 270 days. This delay alone prevents timely benefit adjustment under the offset rules. Finally, BSAS deficiencies have caused substantial delays in automated reconciliation, thereby delaying initial benefit adjustment for some beneficiaries.

8. Overpayments

Several challenges have hindered timely and accurate benefit offset adjustment for BOND treatment subjects (Section 7.5). In this chapter, we document a related outcome: overpayments. We define overpayments, present estimates of BOND's impact on overpayments among T1 subjects during the demonstration's first three years, and present information on beneficiaries' perceptions and reactions to overpayments.

8.1. Definition of Overpayments

Overpayments occur when SSA pays beneficiaries more than they are entitled. In this report, we focus only on work-related overpayments, which are the most prevalent overpayment type (SSA Office of the Inspector General 2015) and the only type of overpayment directly affected by BOND.

Work-related overpayments may occur for several reasons, all of which relate to the timeliness and accuracy of benefit adjustment. Beneficiary failure to report earnings in a timely manner, revised AEEs, inaccurate AEEs,[81] delays in SSA processing of work CDRs, and BSAS errors may all result in overpayments. Both treatment and control subjects may accrue work-related overpayments while in the BPP and EPE, respectively, after beneficiaries have used their three GP months. Some circumstances that generate overpayments are, however, unique to treatment subjects. Specifically, AEEs and BSAS are not relevant to control subjects and hence do not contribute to overpayments for those beneficiaries.

Overpayments fall into two subcategories. The first is overpayments identified after the accounting period. When SSA identifies the overpayment, it requires beneficiaries to repay the owed amount either by check or through withheld future benefits. Beneficiaries have the right to appeal an overpayment, and SSA may agree to set up a repayment plan to mitigate financial hardship.

The second type of overpayment is identified during the annual accounting period and is called an incorrect payment. In these cases, SSA withholds benefit checks immediately until the payment is recovered or until the end of the calendar year, whichever comes first.[82] Control beneficiaries are subject to current law, under which SSA uses a monthly accounting period to adjust benefits and, by definition, identifies overpayments after the end of the monthly accounting period. That is, control subjects may not accrue incorrect payments.

SSA adjusts treatment subjects' benefits based on a calendar-year accounting period, and those beneficiaries may encounter incorrect payments. Given that SSA continues to pay treatment subjects' benefits monthly, each month's benefit is based on average monthly earnings over the entire calendar

[81] To have an accurate AEE, beneficiaries must accurately predict not only earnings but must also accurately predict and account for any non-countable income, such as paid time off and impairment-related work expenses.

[82] At the end of each calendar year, incorrect payments are eligible to be reclassified as overpayments. According to ORDES staff, SSA withholds benefits until a beneficiary with an incorrect payment either submits a new AEE after the start of a new calendar year or until SSA runs an automated reconciliation for the previous year, whichever comes first. In addition, a beneficiary can ask ORDES to process the overpayment earlier by asking for a beneficiary-initiated reconciliation.

year. Beneficiaries may experience difficulty in estimating annual earnings with accuracy, potentially leading to incorrect payments when earnings are underestimated. Submission of a revised AEE provides an opportunity to account for changes in work activity and earnings, but, because the accounting period for determining monthly benefits is annual, changes in earnings as reflected in revised AEEs submitted after January (or after the first offset month, if later) retroactively affect benefits paid in previous months within the same calendar year and can result in incorrect payments.[83]

In the remainder of this chapter, we use overpayments to refer to both types of work-related overpayments: overpayments and incorrect payments.

8.2. Prevalence of Overpayments

On several occasions since the start of the demonstration, we have interviewed BOND staff about the prevalence of overpayments. In this section, we summarize staff perspectives on overpayments, drawing from interviews with staff and providing new evidence from several sources on the prevalence of overpayments. First, we present information from recent interviews with WIC and EWIC counselors and supervisors. Second, we incorporate new information from interviews with a sample of work-oriented T1 beneficiaries.[84] Even though these sources provide useful qualitative information, they do not provide needed information on the prevalence and typical size of overpayments to T1 beneficiaries or on how the experiences of T1 beneficiaries compare with those of C1 beneficiaries.

For the first time, we are able to use SSA administrative data to provide statistics on the prevalence and size of overpayments among T1 beneficiaries and formally test the impact of the benefit offset on the prevalence and size of overpayments. The analysis uses a measure of overpayments we created using SSA administrative data because there are no readily available statistics on overpayments that accrued to BOND beneficiaries during the demonstration period.[85] It is important to note that, in some cases, the estimated overpayment amounts may not align exactly with SSA records of overpayments, but checks on the estimates indicate they are a good approximation, especially in aggregate (Appendix C). For technical reasons, our sample for the overpayment statistics includes only BOND disabled-worker beneficiaries entitled to SSDI solely by their own earnings histories and only for those with data for each month of the calendar year.[86,87] Given these criteria, we exclude approximately 20 percent of all T1 subjects and C1 subjects from the quantitative overpayment analysis.

---

[83] Beneficiaries with underestimated AEEs that are not corrected via revised AEEs will encounter overpayments.

[84] As described in Section 2.1.1, we include responses from WIC and EWIC staff that apply to both Stage 1 (T1) and Stage 2 (T21 and T22) subjects.

[85] For information on construction of the measure, see Section 2.2.2 and Appendix C.

[86] We focus on disabled-worker beneficiaries as a simplification to avoid data limitations that make it problematic to calculate overpayments for dual-entitled and auxiliary beneficiaries. For more information, refer to Appendix C.

[87] Construction of overpayments requires data from the DBAD file. Beneficiaries are included in the DBAD as long as they have a Ledger Account File code reflecting the beneficiary's current payment status. Current payment, benefit suspension, and termination are common categories included in the DBAD. In some instances, such as death, SSA stops populating these status codes such that the beneficiary is no longer included in the DBAD.

It is important to note that the statistics for overpayments are not contemporaneous with the qualitative information collected on overpayments. The analysis in this chapter examines impacts and descriptive statistics for overpayments from 2011 through 2013. We conducted WIC and EWIC staff interviews in fall 2013 and fall 2014 and conducted interviews with work-oriented beneficiaries in fall 2015. In the sections that follow, we present information based on the reference period, and, when several sources of information were available, we synthesized information from all sources.

8.2.1. Overpayments During the Early Demonstration Years

According to BOND staff, benefit adjustment early in the demonstration was often coupled with overpayments, as SSA and BOND staff were establishing new processes, beneficiaries were learning about their reporting requirements, and SSA was contending with a large backlog of work CDR cases (Gubits et al. 2013, Derr et al. 2015). Indeed, in 2013, WIC and EWIC staff in 7 of the 10 BOND sites said that all or nearly all of the beneficiaries they served who used the offset had had an overpayment (Derr et al. 2015).

Statistics on overpayments support the reports provided by BOND staff. In particular, offset users commonly had overpayments in the first three years of the demonstration (Exhibit 8-1). According to data extracted in October 2015, 83.3 percent of T1 subjects who used the offset in 2011, 2012, or 2013 had overpayments that accrued during that period. Given that the fraction of T1 subjects using the offset during the period is small (BTS data from October 2015 indicate that 2.3 percent of T1 subjects had used the offset in this period), the fraction of all T1 subjects with an overpayment is also small, at 2.1 percent.

The prevalence of overpayments varied across the demonstration's first three years. Both the proportion of all T1 subjects overpaid and the proportion of T1 offset users who were overpaid were lowest in 2011, perhaps because (1) we analyzed 8 months of overpayments in 2011 (May—the first month in which T1 subjects could use the offset—through December 2011) versus 12 months in 2012 and 2013 and (2) T1 beneficiaries in EPE suspense before random assignment would have had underpayments rather than overpayments following any delay in benefit adjustment at the start of the demonstration.[88] It is also important to note that the proportion with an overpayment during this period—particularly in more recent years—will likely increase as SSA receives and processes new information on beneficiary earnings.[89]

[88] Underpayments occur when beneficiaries receive less in benefits than they were entitled (Section 8.4). Under current law, beneficiaries in the EPE who engage in SGA are not entitled to receive cash DI benefits, while treatment subjects with the same earnings may be entitled for a partial benefit under the BOND offset. Treatment subjects who were in EPE suspense before BOND and are entitled to a partial benefit while in BOND are underpaid if their benefits are not adjusted timely under the offset rules.

[89] The data presented in the report are based on data available in the DBAD through October 2015. We reproduced statistics for 2011 and 2012 by using data from April 2014 DBAD. When using the additional 18 months of data, the rate of identified overpayments among all T1 subjects increased from 0.64 to 0.77 percent in 2011 and from 0.85 to 1.21 percent in 2012. There was also an increase for C1 subjects from 0.92 to 0.96 percent for 2011 and from 0.98 to 1.05 percent for 2012. For both groups, the increase was larger in absolute and relative terms for the 2012 statistics compared to the 2011 statistics.

Given that the prevalence of overpayments among all T1 subjects is small, the mean overpayments are likewise small during each period (column 3 of Exhibit 8-1), but the mean for those with overpayments is much larger (column 4). The mean T1 overpayment during the 32-month analysis period was $122 (in 2011 dollars), reflecting the implicit averaging of $0 overpayments across a large majority of the sample. For those with an overpayment during any month of the 32-month period examined, the mean overpayment for the same period is almost $5,800. T1 subjects with any overpayment during this period were overpaid for an average of almost 13 months. Though large, the mean *monthly* overpayment of $416 in months with any overpayment (not shown) is more than one-third of the typical monthly SSDI benefit payment in December 2013, at $1,130 (SSA 2013a), and reflects the partial-benefit payments provided under the benefit offset. That is, for a given duration of overpayment, earnings amount, and size of the full benefit amount, treatment subjects receive smaller overpayments than do current-law beneficiaries because of the offset.

Exhibit 8-1. Prevalence of Overpayments in 2011, 2012, or 2013

Period	T1 Subjects with Overpayment (%)	T1 Offset Users with Overpayment (%)	Mean Overpayment in Period for All T1 Subjects[a]	Mean Overpayment in Period for T1 Subjects with Overpayment in Period[a]
May–December 2011	0.77	62.9	$21	$2,766
January–December 2012	1.21	72.4	$49	$4,080
January–December 2013	1.34	69.8	$51	$3,835
May 2011–December 2013	2.11	83.3	$122	$5,786

Source: DBAD extracts from May 2011–December 2013 and October 2015.

Note: Values are not regression-adjusted. We estimate overpayments starting in May 2011, the first month following BOND random assignment and the first month T1 subjects could use the offset.

[a] We used the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W) to adjust 2012 and 2013 dollars for inflation to be equivalent to 2011 dollars.

Unweighted sample sizes: T1 (2011) = 65,234, T1 (2012) = 65,337, T1 (2013) = 65,338

BOND's experimental design supports a rigorous (but exploratory) analysis of the impact of the benefit offset on the rate and size of overpayments among all T1 subjects. The analysis reveals that the prevalence of overpayments was larger among all T1 subjects than among all C1 subjects during BOND's second and third years, but smaller in the demonstration's first year (Exhibit 8-2). The 2011 and 2013 estimates provide strong evidence (that is, were significant at the 1 percent level) that the benefit offset had an effect on overpayments (in opposing directions), while the 2012 estimate provides evidence (that is, significance at the 5 percent level) of an effect. The percentage of T1 subjects with an overpayment across the entire period is 0.23 percentage points higher than for C1 subjects, or 12.2 percent of the control group percentage. In 2011, the reverse is true; the percentage of T1 subjects with an overpayment is 0.15 percentage points lower than the percentage among C1 subjects in that year. The difference may be an artifact of the transition from current-law rules to BOND rules, reflecting T1 subjects who were in benefit suspense during the EPE, as already noted.

Exhibit 8-2. Estimated Impacts on Overpayments in 2011, 2012, and 2013

Outcome	T1 Mean	C1 Mean	Impact Estimate
Percentage with Overpayment			
Overpaid in any month in 2011 (%)	0.77	0.92	-0.15*** (0.04)
Overpaid in any month in 2012 (%)	1.21	1.07	0.14** (0.05)
Overpaid in any month in 2013 (%)	1.34	0.99	0.35*** (0.06)
Overpaid in any month in 2011, 2012, or 2013 (%)	2.11	1.88	0.23*** (0.07)
Mean Overpayment Amount[a]			
Mean 2011 overpayment	$21	$49	-$28*** ($3)
Mean 2012 overpayment	$49	$71	-$22*** ($3)
Mean 2013 overpayment	$51	$67	-$15*** ($3)
Mean combined 2011, 2012, and 2013 overpayment	$122	$187	-$65*** ($7)

Source: DBAD extracts from May 2011–December 2013 and October 2015 and baseline SSA administrative data.

Note: We estimate overpayments only for months following BOND random assignment. All comparisons are inclusive of all T1 subjects and all C1 subjects.

[a] We used the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W) to adjust 2012 and 2013 dollars for inflation to be equivalent to 2011 dollars.

Unweighted sample sizes: T1 (2011) = 65,234, T1 (2012) = 65,337, T1 (2013) = 65,338, C1 (2011) = 718,839, C1 (2012) = 719,301, C1 (2013) = 719,289

*/**/*** Impact difference is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test.

Several factors may have increased the prevalence of overpayments under the offset relative to current law. The first factor is the delays in processing benefit adjustments under the offset because of resource constraints and problems with the adjustment process (Chapter 7). Second, the switch from a monthly accounting period under current law to an annual accounting period under the BOND offset might increase the incidence of overpayments—particularly small overpayments—because of discrepancies between predicted annual earnings (on AEEs) and actual earnings. Finally, T1 subjects who begin work in months after the start of their BPP (in the first year of the BPP) or the start of the calendar year (for subsequent years) may be subject to incorrect payments. The second and third factors both relate to the annual accounting period under BOND and thus are not applicable to C1 subjects.

Although T1 subjects were more likely than C1 subjects to have an overpayment, we find strong evidence that T1 subjects had lower overpayment amounts than did C1 subjects in each of the first three years of the demonstration (Exhibit 8-2). The $65 reduction in mean overpayments over the entire period represents 35 percent of the C1 mean. Two factors determine the direction and size of the impact: the relative prevalence of overpayments and the relative size of overpayments among those overpaid. Given that T1 subjects were more likely to have overpayments than C1 subjects, the latter effect must have dominated. That is, in months with overpayments, C1 subjects were overpaid significantly more than T1 subjects. Indeed, for beneficiaries with a given earnings amount and size of the full benefit, the monthly

T1 overpayment amount may be no larger than the monthly C1 overpayment amount and is typically much smaller.

For T1 subjects with an overpayment, the effect of assignment to the BOND treatment group on the mean size of overpayments is presumably much larger than the effect on the mean overpayment for all T1 subjects. The size of the effect on mean overpayments for all T1 subjects is not especially large because it averages in the zero impact for the 97.9 percent of T1 subjects who did not have an overpayment during the demonstration's first three years. We cannot calculate the true impact among the subset of overpaid beneficiaries, because we cannot identify the counterparts of the T1 subjects with overpayments among the C1 subjects. Instead, we make an informal calculation as an approximation, based on what would seem a reasonable, but unverifiable assumption. We assume that all C1 subjects with an overpayment would have had an overpayment if they had been assigned to T1. Under that assumption, the difference in the effect of the offset on the mean overpayment among T1 subjects with overpayments during any period is equal to the mean impact for all T1s divided by the percentage of T1 subjects with an overpayment during the same period. The calculation produces mean differences for T1 subjects with overpayments of about -$3,600, -$1,800, and -$1,100 in 2011, 2012, and 2013, respectively, and of -$3,100 over all three years for those with overpayments in any month in the period.[90] The decline in the numbers over the three years reflects the increase in the prevalence of overpayments among T1 subjects. The assumption required to produce these estimates may not be exactly correct—perhaps some C1 subjects would have avoided overpayments altogether if they had been assigned to T1—but the assumption seems unlikely to be so substantially violated that the order of magnitude of the estimates is misleading.

8.2.2. Overpayments During Later Demonstration Years

According to BOND staff interviewed in fall 2014, overpayments remained common among beneficiaries with first offset adjustments as of that time. Implementation Team staff attributed overpayments to a combination of beneficiaries failing to report earnings and delays in work CDR processing. In either case, the result is a delay in benefit adjustment. Statistics from 2014 and 2015 reveal a median interval of nearly two years between when a beneficiary first used the offset and when SSA made the initial adjustment (Section 7.5.2). WIC and EWIC staff reported that it seems almost inevitable that offset users would have an overpayment—especially an incorrect payment—during the first year in which SSA adjusted their benefits. Implementation Team staff responsible for post-entitlement work agreed that it was rare for a beneficiary not to have an overpayment when SSA made the initial benefit adjustment.

According to WIC and EWIC staff interviewed in 2014, overpayments were generally less frequent and smaller following the first offset adjustment. This is in-part by design, because offset users are required to submit AEEs for future calendar years to avoid benefit suspension. Reports from both ORDES and Implementation Team staff suggest that improvements in the accuracy of AEEs have helped to facilitate proper adjustments.

[90] Because the assumption that C1 subjects who were overpaid also would have been overpaid had they been assigned to the BOND T1 group might not be exactly correct, this should be considered an upper bound for the point estimate of the mean reduction in the size overpayments for those who would have an overpayment under the BOND benefit offset, current law, or both.

8.3. Beneficiary Experiences with Overpayments

The 20 T1 beneficiaries interviewed in 2015 who had used the offset in a least one year experienced overpayments consistent with reports from BOND staff and descriptive statistics on overpayments. Of these 20 interviewees, 15 reported overpayments.[91] Twelve beneficiaries with overpayments reported that their overpayments occurred while in BOND, 2 reported overpayments that predated BOND, and one did not provide information about the timing of the overpayment. In addition, 3 of the 10 work-oriented T1 beneficiaries who had not used the offset also reported overpayments that predated BOND, again highlighting that overpayments are not unique to BOND. In total, 18 of the 30 work-oriented T1 subjects we interviewed reported past overpayments.

Earlier reports have documented beneficiary experiences with overpayments, including the overpayments' role in producing financial hardship (Derr et al. 2015) and creating negative perceptions of the demonstration and its staff (Gubits et al. 2013). During interviews in 2014, all participating WIC and EWIC staff indicated that the beneficiaries they serve generally have negative reactions to overpayments. About one in three staff observed that, for the most part, beneficiaries subsequently recover from the initial negative reaction.

The work-oriented beneficiaries included in our in-depth interviews described a mixed range of reactions to overpayments. Several gave neutral responses to overpayments; for example, one offset user who experienced overpayments both before and during BOND said that the overpayments did not change her perception of the program because she knew it was just a problem with SSA. Another noted that overpayments are "just part of the process." A third beneficiary observed that BOND and SSA staff were looking out for his best interests.

In contrast, some beneficiaries described negative reactions and adverse financial outcomes following overpayments. One-quarter of interviewed offset users highlighted overpayments as a negative feature of BOND. One beneficiary said that receipt of an overpayment made her "bitter and angry" while another said that she felt "like I was being punished for working." A third offset user described significant financial struggles that she attributed to a one-year delay in SSA processing her submitted paperwork, with a resulting incorrect payment. Another beneficiary said she had to borrow money because she could not otherwise fulfill her financial obligations without the benefit check she was expecting when overpayments caused it to be withheld. It is important to note that, although these beneficiaries clearly and firmly attributed their financial struggles to overpayments, there may have been other causes. Further, the situations they described could also have occurred, or even have been worse, under current law.

Beneficiary reactions to overpayments varied by duration of offset use. Among the 15 work-oriented interviewees who had used the offset, long-term offset users were less likely to see overpayments as reflecting poorly on BOND (1 of the 9 long-term offset users with overpayments). Short-term offset users were more likely to experience overpayments as a negative reflection on BOND (3 of the 6 short-term

[91] Beneficiaries may receive overpayments for a variety of reasons, and many beneficiaries were not sure of or did not mention the reason for their overpayment. However, because all beneficiaries in our sample are eligible for work-related overpayments, they are the most prevalent type of overpayment (SSA Office of the Inspector General 2015), and many other overpayment reasons are not germane to our interview sample (e.g., payment issued after death, medical improvement), we assume that beneficiaries referred to work-related overpayments.

offset users with overpayments) and were more likely to describe serious inconveniences attributable to these payment issues (3 of the 6 short-term offset users with overpayments, 1 of the 9 long-term offset users with overpayments).

Some beneficiaries may have changed their employment decisions because of overpayments. In our poll of 49 WIC and EWIC staff, 43 percent of respondents reported that they served a beneficiary who reduced earnings in response to an overpayment. Implementation Team staff heard anecdotal reports of similar beneficiary responses to overpayments. One of the 18 beneficiaries with overpayments who participated in our in-depth interviews reported that she decreased her earnings after receiving notice that her benefit would stop because she owed SSA money. She explained, "Why try to go back to work if they're going to penalize me for trying." It is difficult to know, however, whether the beneficiary was responding to the overpayment per se or to a new realization that she could not continue to earn above SGA and keep her full benefits. The findings from beneficiary interviews in combination with the poll results may mean that such responses to overpayments are relatively rare; nonetheless, the responses remain well known among counselors who have encountered many clients with overpayments.

Deliberate earnings reductions may reflect several factors. We would expect such earnings reductions to be relatively infrequent because the result is a reduction in total income. Beneficiaries, however, (1) may not understand that reductions in earnings could lead to a reduction in total income, (2) may not believe that the BOND offset rules will be applied as intended, or (3) may decide that they are better off working fewer hours despite the loss of income. It is likely that beneficiaries subject to current law exhibit similar responses to overpayments, but, in their case, reducing their earnings to below the SGA level might increase their income.

One of the 18 interviewees with overpayments had the opposite response to encountering an overpayment. He increased his hours of work each week because his benefit was withheld for three months due to an overpayment. He explained that he had to "make up the difference" in order to be able to pay his bills.

BOND staff suggested that negative reactions to overpayments may reflect in part the nature and method of SSA communications. WIC and EWIC staff reported that beneficiaries often note that SSA's notices of overpayments are confusing. Implementation Team staff added that the notices are alarming, although the same is true for overpayment notices received by all SSDI beneficiaries. The fact that WIC and EWIC staff do not receive copies of the notices means that they sometimes have difficulty helping their clients understand the implications of the notices; the cognizant counselor must ask the client either to send him or her a copy of the notice or read it over the telephone. If the beneficiary does not disclose the information contained in a notice, WIC and EWIC staff must necessarily base their beneficiary counseling activities on incomplete information. Under current law, the same challenge applies for beneficiaries and WIPA-funded counselors. ORDES staff noted that SSA chose not to provide direct access to overpayment notices to the BOND implementation team because of workload constraints in the SSA BOND work unit and to be consistent with the standard for WIPA-funded counseling. WIC, EWIC, and WIPA staff may counsel beneficiaries about overpayments, but the SSA field office is responsible for assisting all SSDI beneficiaries with overpayments (for example, setting up repayment plans). However, members of the Implementation Team reported that, even though SSA field offices are obligated to assist T1 beneficiaries with overpayments, some field office staff members are unclear about their role and may decline to provide that service.

WIC and EWIC staff may help mitigate the potential negative effects of overpayments on beneficiary behavior. According to WIC and EWIC staff, some beneficiaries plan their expenditures to minimize financial difficulties if they are aware of pending overpayments (Derr et al. 2015). Over the course of the demonstration, WIC and EWIC counselors have become more adept in anticipating overpayments and counseling beneficiaries on the topic. In 2012, for example, WIC and EWIC staff struggled to identify and notify beneficiaries of pending overpayments (Gubits et al. 2013). In 2013, WIC and EWIC staff reported that they were generally able to alert beneficiaries to a potential overpayment (Derr et al. 2015). Indeed, relative to previous years, BOND staff reported that in 2014 they were better able than in years past to help treatment subjects anticipate overpayments and the effects on beneficiary incomes. Implementation Team staff explained that additional training and centralization of post-entitlement work helped improve the process. In addition, guidance that directs staff discuss overpayments whenever they collect an AEE appears to have helped beneficiaries. However, during the 2014 interviews, Implementation Team staff noted that overpayments were a lingering source of confusion for some WIC and EWIC staff.

WIC and EWIC staff may also help beneficiaries reduce the likelihood of overpayments. Reflecting perceptions about beneficiary aversion to overpayments, WIC and EWIC staff reported that they generally encourage beneficiaries to submit AEEs that are slightly higher than earnings expectations for the coming year so that the beneficiaries will be less likely to receive overpayments after automated reconciliation.

8.4. Underpayments

The previous discussion has not considered the converse of overpayments: underpayments. Underpayments occur when beneficiaries receive less in benefits than they were entitled. When SSA recognizes an underpayment, it issues beneficiaries a lump-sum check. Refer to the *Process Study Report* for additional discussion of underpayments.

The exact rate of underpayments among BOND subjects is unknown.[92] The perception among ORDES staff in early 2015 was that, after the first benefit adjustment under the offset, underpayments are at least as likely to occur as overpayments. However, only 2 of the 20 offset users among the T1 respondents to our in-depth interviews in 2015 reported having an underpayment; the same 2 offset users also reported having an overpayment. It is possible that underpayments are not as salient to beneficiaries as are overpayments and therefore go underreported. It is also possible that the reports of underpayment are an artifact of conducting in-depth interviews with a sample that is not necessarily representative of all offset users.

[92] There are no readily available statistics on underpayment. To identify work-related underpayments would require a distinct algorithm. Consistent with the *Evaluation Analysis Plan*, this analysis focuses on overpayments.

8.5. Summary

Overpayments are prevalent among T1 offset users, particularly when SSA first adjusts benefits under the offset. According to October 2015 SSA administrative data, in the demonstration's first three years, more than 83 percent of T1 offset users had an overpayment. The mean amount of overpayments that accrued over the three years totaled $5,917. In 2014, the year after the three years covered by these data, WIC and EWIC staff observed that beneficiaries were still highly likely to have an overpayment during their first year of benefit adjustment but that overpayments were less likely and smaller in subsequent years of offset use.

We find strong evidence from an exploratory analysis that assignment to the BOND treatment group increased the likelihood of an overpayment in the demonstration's first three years. The average overpayment in that period for the full T1 group (including $0 values for those not receiving overpayments), however, was smaller than for the full C1 group. During the demonstration's first three years, T1 subjects had a 0.22 percentage point higher likelihood of an overpayment than C1 subjects, a 12 percent increase over the C1 mean. Over the same period, T1 subjects accrued $65 less in mean overpayments relative to C1 subjects, a difference that is 35 percent of the C1 mean. When the reduction in overpayments is spread over the 2.1 percent of T1 subjects with overpayments, the mean reduction over the three-year period totals about $3,100.

In in-depth interviews, beneficiaries with overpayments exhibited a range of responses. Some T1 beneficiaries had neutral reactions to overpayments while others formed negative associations between overpayments and the BOND offset, and two attributed substantial financial difficulties to the overpayments. According to some reports, beneficiaries reduced their earnings after an overpayment, but this appears to have occurred in only a small minority of cases. In addition, one beneficiary reported increasing earnings in response to an overpayment.

Although BOND staff attempt to mitigate the effects of overpayments, their efforts have no effect on the payment outcome, which is largely influenced by the time frames of SSA's information gathering and processing efforts. Indeed, the 22-month average delay between offset use and initial benefit adjustment represents 22 months during which SSA may have overpaid beneficiaries.

9. Impacts on Annual Earnings and SSDI Benefits Measured in Administrative Data

This chapter presents estimates of the impact of the BOND benefit offset on the annual earnings and disability benefits of SSDI beneficiaries in 2014. We estimate the impact of the offset with standard work incentives counseling relative to current law (T1 versus C1). Those randomly assigned to T1 became subject to SSDI benefit payments under the offset starting in May 2011. Hence, the duration of subjects' participation in Stage 1 at the end of 2014 was 44 months. Later reports will examine impacts in 2015 and beyond when all treatment group subjects have been subject to the BOND benefit payment rules for longer periods.

All the findings in this chapter involve outcome measures taken from SSA administrative data; impact results from the Stage 1 36-Month Survey appear in Chapter 10. We also refer the reader to Chapter 2 for definitions of the outcome variables, theories about possible impacts, administrative features of the offset that may influence impacts, and the impact estimation methodology used to generate the results presented in this chapter. We organize the chapter into four sections. Section 9.1 provides confirmatory impact evidence on annual earnings and total SSDI benefits paid in 2014. Section 9.2 presents exploratory evidence regarding other employment- and benefit-related outcomes while Section 9.3 highlights variation in earnings and benefit impacts by beneficiary background characteristics. Section 9.4 summarizes the chapter's findings.

When discussing the impact estimates presented in this chapter, we use particular language to signify different levels of confidence that a non-zero impact has occurred, as defined in Chapter 2. We classify results with 0.01, 0.05, and 0.10 levels of statistical significance as providing strong evidence, providing evidence, and providing some evidence that the offset had an effect on the tested outcome, respectively.

We make a distinction between hypothesis testing of a confirmatory versus exploratory nature. Statistically significant findings from our predesignated *confirmatory* analyses meet a higher standard of evidence—one that minimizes the possibility of "false positive" findings (i.e., apparent impacts where the true impact is zero) by adjusting the *p*-values of the tests. In contrast, statistically significant findings from *exploratory* hypothesis tests offer suggestive evidence of other impacts that the benefit offset may have achieved. Confirmatory tests are limited to results on total earnings and total SSDI benefits paid in 2014. All results generalize statistically to represent the national population of SSDI beneficiaries.

In brief, the impact estimates provide strong evidence of an increase in total SSDI benefits paid in 2014, but no evidence of an impact on 2014 total earnings. For the exploratory outcomes, we find evidence of a positive impact on both employment and earnings above BYA, some evidence of a negative impact on earnings above three times BYA, and strong evidence of a positive impact on number of months with SSDI payments during 2014. The analysis of subgroup impacts by beneficiary background characteristics provides some evidence of effect variation by duration of earlier benefit receipt, 2010 employment status, beneficiary age, and type of disability (back disorders versus other diagnoses). All statistically significant impacts are in the direction predicted by theory, in cases where theory offers a clear prediction (Chapter 2).

9.1. Confirmatory Impacts on 2014 Earnings and SSDI Benefits

This section presents impact estimates for the beneficiary outcomes of paramount policy interest for the demonstration: total earnings and total SSDI benefits paid in 2014, which is the year with the most recently available data.[93] We selected the outcomes solely on the basis of theory and policy interest before outcome data collection (*Final Design Report* and *Evaluation Analysis Plan*). For the outcomes, we examine whether the benefit offset—as a package, including both the benefit offset and associated administrative processes—has an impact on beneficiaries as compared to current law. We also present and discuss annual trends in these outcomes since Stage 1 random assignment in May 2011. For both outcomes, the sign of the theoretical prediction for impacts is ambiguous.

The impact analysis in this section involves two hypothesis tests: one for each of the two outcomes comparing the difference between the T1 and C1 means. We perform a multiple-comparison adjustment procedure on the two tests together so that the *p*-values account for the higher chance of a Type 1 error—the probability of rejecting at least one null hypothesis in a family of hypothesis tests when all null hypotheses are true.[94] We carry out the adjustment for these two confirmatory outcomes because they are of paramount interest to the BOND. All other impact findings reported in this chapter are considered exploratory; therefore, we made no multiple-comparison adjustments for them.

For total earnings received from January through December 2014, we find no evidence of an effect on treatment group subjects relative to control group members (first row of Exhibit 9-1). The point estimate for this impact is $17 and has a *p*-value of 0.550, which is well above the standard for statistical significance ($p < 0.10$).[95]

[93] These two outcomes were identified in the *Evaluation Analysis Plan* for confirmatory analysis before the research team had access to outcome data for study subjects. Prespecifying outcomes for confirmatory analysis before gaining access to outcome data is standard scientific practice and avoids the possibility that researchers will select data that best support a particular type of policy conclusion. (See the discussion of confirmatory outcomes in Section 6.1 of the *Evaluation Analysis Plan*.) In later reports, we will supplement the impacts on earnings and SSDI benefits for 2014 as confirmatory findings with impact estimates for the same outcomes in subsequent years. The practice of replacing previous confirmatory impact estimates with the most recently available estimates reflects the supremacy of long-term impacts in determining an intervention's overall effectiveness in achieving its long-term goals.

[94] See *Third-Year Snapshot of Earnings and Benefit Impacts for Stage 1* for more details about the multiple-comparisons adjustment procedure.

[95] The evidence of impact on total earnings in 2014 is not statistically significant even when the multiple-comparisons adjustment is not made to the *p*-value.

Exhibit 9-1. Estimated Impacts on Total Earnings in 2014 and Total SSDI Benefits Paid for 2014

Outcome	T1 Mean	C1 Mean	Impact Estimate
Total earnings [a] (January–December 2014)	$1,385	$1,368	$17[a] ($27)
Total SSDI benefits paid (January–December 2014)	$11,230	$11,098	$132[b]### ($28)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data (used as covariates in impact analysis regression equations).

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 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. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: T1 = 77,115, C2 = 891,598

#/##/### 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 0.550 and hence does not provide confirmatory evidence of an impact.

[b] The impact estimate for total SSDI benefits paid has a *p*-value after multiple-comparison adjustments of 0.002 and hence provides confirmatory evidence of an impact.

The estimate of the benefit offset's impact on 2014 earnings mirrors the impact findings for earnings impacts from the demonstration's previous years (Exhibit 9-2). Unlike Exhibit 9-1, the dollar amounts in Exhibit 9-2 have been adjusted for inflation in order to make equivalent comparisons across years. Point estimates in each year are positive, but none is significant. Over the four-year demonstration period, point estimates for average annual earnings (in 2011 dollars) have increased for each of the Stage 1 random assignment groups taken individually—from $1,193 to $1,319 for T1 subjects (an 11 percent increase) and from $1,201 to $1,302 for C1 subjects (an 8 percent increase).[96]

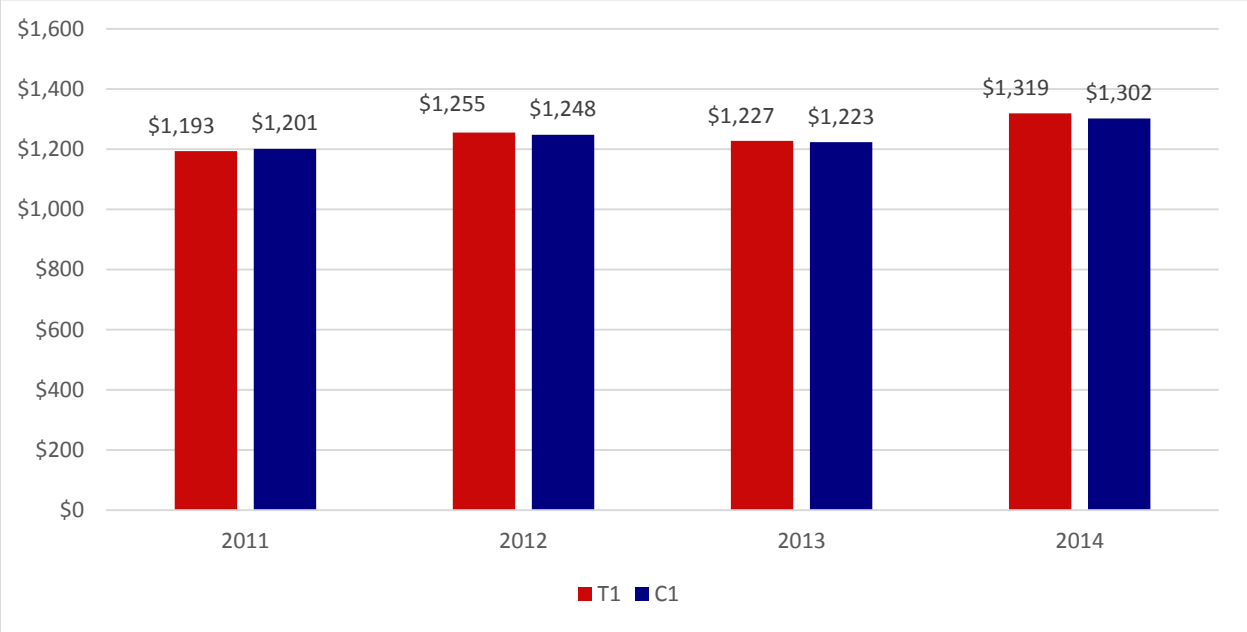
The impact estimates provide strong evidence of a positive effect of the demonstration on total SSDI benefits paid in 2014 (January to December; second row of Exhibit 9-1). SSDI benefits paid to C1 subjects under current law are $11,098 per year; SSDI benefits paid to T1 subjects under the offset are $11,230. The difference of $132 is equivalent to one percent of benefits under current law and is statistically significant (adjusted *p*-value of 0.002).

This finding mirrors the evidence on SSDI benefits from earlier analyses. All four demonstration years to date point to strong evidence that the benefit offset increases SSDI benefits paid. Given that the 2011 estimate pertains to only eight months, it is helpful to make a rough (ignoring any seasonality) conversion of annual impact measures to monthly values, dividing by the number of months in the year (Exhibit 9-3).

[96] Although we have not conducted a formal test, its seems likely that these changes are statistically significant, given the size of the standard errors for differences in earnings between T1 and C1 within a year and the panel nature of the data, which is likely to reduce standard errors for within-group differences in earnings across years.

Average monthly SSDI benefits paid (in 2011 dollars) to the individual random assignment groups declined slightly each year, from $941 to $891 between 2011 and 2014 for T1 subjects (a 5 percent decrease) and from $938 to $880 for C1 subjects (a 6 percent decrease). The decline in mean benefits paid to T1 and C1 subjects results from SSA suspending or reducing SSDI benefits for subjects in both groups. Reasons for suspensions and terminations include mortality, medical improvement, and benefit reductions due to work.

Exhibit 9-2. Trends in Annual Earnings (in 2011 Dollars) in the First Four Years of BOND

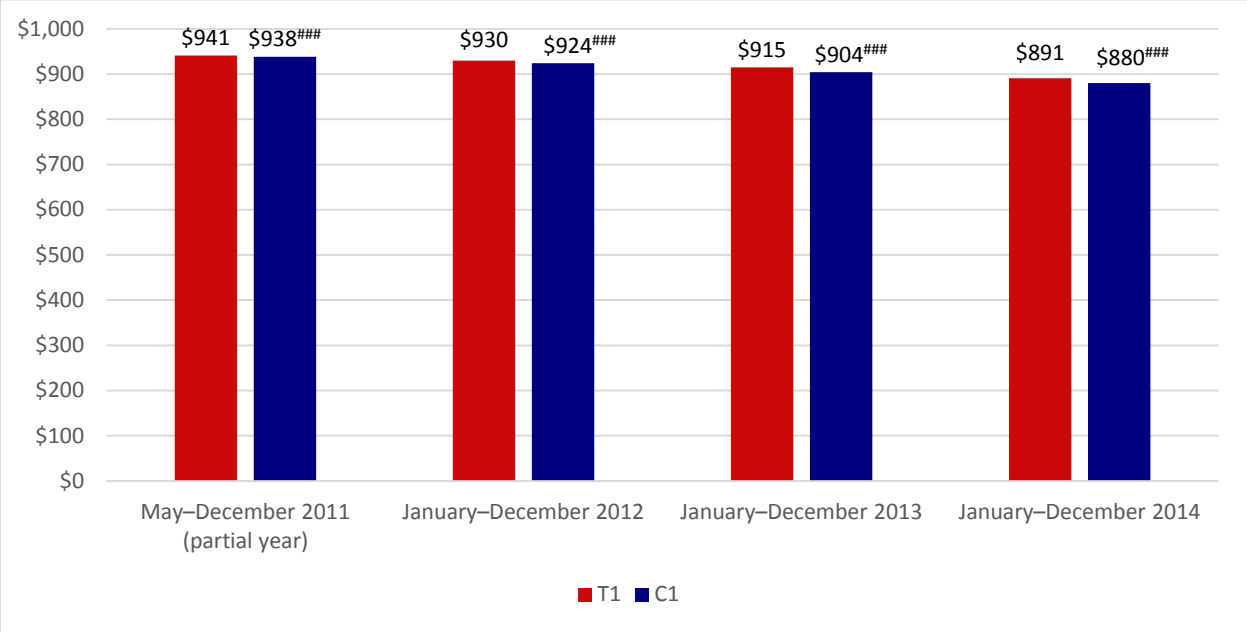


Source: SSA administrative records for calendar years 2011-2014 and baseline SSA administrative data (used as covariates in impact analysis regression equations). See Exhibits A-3, A-6, and A-9 for estimates for 2011, 2012, and 2013 respectively.

Notes: We used the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W) to adjust 2012, 2013, and 2014 dollars for inflation to be equivalent to 2011 dollars. Weights ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment. Means are regression-adjusted for baseline characteristics. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail).

#/##/### 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). For each year, the difference between C1 and T1 means was not statistically significant based on a confirmatory standard of evidence and a two-tailed t-test with 9 degrees of freedom.

Exhibit 9-3. Trends in Average Monthly SSDI Benefits Paid (in 2011 Dollars) in the First Four Years of BOND



Source: SSA administrative records for calendar years 2011-2014 and baseline SSA administrative data (used as covariates in impact analysis regression equations). See Exhibits A-3, A-6, and A-9 for estimates for 2011, 2012, and 2013 respectively.

Notes: See Chapter 2 for variable definitions. We used the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W) to adjust 2012, 2013, and 2014 dollars for inflation to be equivalent to 2011 dollars. Weights ensure that the BOND subjects who met analysis criteria are representative of the national beneficiary population in the month of random assignment. Means are regression-adjusted for baseline characteristics.

#/##/### 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).

The point estimates for impacts on benefits paid increased from 2011 through 2013, but did not increase again in 2014. On a monthly basis, the benefit impact estimates for 2011 through 2014 (in 2011 dollars) are $3.00, $5.71, $11.11, and $10.47, respectively. Although we do not formally test to determine whether the differences across years are statistically significant, it is apparent that the estimates for 2013 and 2014 are significantly different from the point estimate for 2011, and might also be significantly different from the point estimate for 2012.[97]

[97] The assessment of statistically significance is based on the following information. For 2014, the point estimate of the standard error for the impact on mean monthly benefits paid in each year is $2.33 ($28/12 = $2.33). The corresponding value for 2011 is $1.25 ($10/8), based on Exhibit 3-1 of the *First-Year Snapshot Report*. If the samples for the two years were independent, the standard error for the difference between the 2014 and 2011 estimates would be 2.64 [$= (2.33^2 + 1.25^2)^{1/2}$], and the t-statistic for the test of the null hypothesis of no difference would be 2.83 [= (10.47 – 3.00)/2.64)], which has a p-value of less than 0.01. The fact that these are panel data, rather than independent samples, reduces the true standard error for the cross-year differences in impacts relative to what it would be if the samples were independent, as assumed in the above calculation.

9.2. Exploratory Impacts on Other Earnings and Benefit Outcomes

In addition to the two confirmatory outcomes discussed in Section 9.1, we use administrative data to estimate impacts on seven other earnings and benefit outcomes: any employment during 2014 and in various dollar ranges of 2014 earnings relative to BYA, number of months of SSDI receipt during 2014, and total dollar amount and number of months of SSI payments that year. We report impact estimates for these outcomes in this section.

Consistent with the *Evaluation Analysis Plan*, we consider all the analyses in this section to be exploratory and therefore do not make any correction for multiple comparisons. As a result, any statistically significant findings represent suggestive evidence of areas in which further effects of the benefit offset may have occurred; we are not as confident of these results as of the confirmatory impact findings presented above. Even if the offset had no impact on any of the measures examined in this section, there is a greater probability for some impact estimates to be statistically significant solely by chance (relative to the confirmatory results) given that we conducted many hypothesis tests without adjusting for multiple comparisons.

9.2.1. Estimated Impacts on Earnings-Related Outcomes

Even though the confirmatory analysis does not provide evidence of an impact on earnings averaged across all T1 subjects, effects are still possible on either the share of those subjects with any earnings (i.e., non-zero earnings) in 2014 or the share of those with earnings in particular dollar ranges relative to BYA. As stated in Chapter 2, theory does not make a clear prediction for total earnings, but it does predict a positive effect on the percentage with employment and on the proportion of beneficiaries earning above BYA. Both of these effects could occur without the benefit offset affecting earnings on average across the whole sample. In general, the offset is predicted to have dual effects on earnings: a *positive* effect on average earnings for those who would earn below SGA under current law (that is, without the offset) and a *negative* effect on average earnings for those who would earn above SGA under current law. It is important to note that the two groups are not directly observed in the data.

The evidence suggests that the benefit offset had an impact on three of the four exploratory employment measures in 2014, in directions consistent with theoretical predictions (panel one of Exhibit 9-4). The impact estimates provide evidence that the offset: increased the 2014 employment rate among T1 subjects; increased the share of beneficiaries earning above BYA in 2014; and reduced the share of beneficiaries earning more than three times BYA, with the evidence on the first and last impact weaker by our usual standard than for the second impact. Theory does not, however, predict a particular threshold level of earnings above which reductions from current law earnings levels might occur. There is no statistically significant impact in 2014 on the share of beneficiaries with earnings greater than two times BYA.

These findings are consistent with those for previous years. For example, the point estimates of effects on each of the exploratory earnings-related outcomes in 2013 had the same sign as the corresponding estimate in 2014, but none was statistically significant (Wittenburg et al. 2015, Exhibit 3-1).

The exploratory earnings outcome findings suggest a partial explanation of why the estimated impact on total earnings is not significant: positive impacts on earnings for those at the low end of the earnings distribution under current law are at least partially offset by negative impacts for those at the high end of

the current-law earnings distribution. At the low end of the current-law earnings distribution, under the offset about 0.20 percent more beneficiaries—about one in 500—have earnings above BYA, relative to what would have been expected under current law. At the high end of the earning distribution, under the offset, 0.07 percent fewer—about one in 1,400—have earnings above three times BYA relative to what would have been expected under current law. All but one of these coefficients are significant at the 10 percent level. The resulting net impact on total earnings, if any, is not large enough to be distinguished from sampling error, whatever its sign.

Exhibit 9-4. Estimated Impacts on 2014 Employment and Benefit Receipt

Outcome	T1 Mean	C1 Mean	Impact Estimate
Employment (January–December 2014)			
Employment during year (%)	13.18	12.90	0.28* (0.13)
Earnings above BYA (%)	2.89	2.69	0.20** (0.07)
Earnings above 2 times BYA (%)	1.23	1.28	-0.05 (0.04)
Earnings above 3 times BYA (%)	0.64	0.71	-0.07* (0.03)
Benefit Receipt (January–December 2014)			
Number of months with SSDI payments	10.50	10.39	0.10*** (0.02)
Total SSI benefits paid	$417	$422	-$5 ($8)
Number of months with SSI payments	1.81	1.82	-0.02 (0.01)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data (used as covariates in impact analysis regression equations).

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 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. Dollar values are not adjusted for inflation.

Unweighted sample sizes: T1 = 77,115, C2 = 891,598

*/**/*** 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.

It is worth noting that there is a substantial increase in the percentage with earnings in the earnings range just above the earnings cliff (BYA to two times BYA) because theory predicts that this earnings range is highly unattractive under current law; to earn in this range, beneficiaries would have to give up monthly benefits that are in most cases in excess of BYA, whereas they can keep all of their benefits if they keep their earnings just below BYA. The percentage of the treatment group with earnings in this range is 0.25 percentage points higher (1.66 = 2.89 – 1.23) than for the control group (1.41 = 2.69 – 1.28), an 18 percent increase. Consistent with theory, it appears that the benefit offset makes this earnings bracket more attractive to beneficiaries.

9.2.2. Estimated Impacts on Benefit-Related Outcomes

The impact estimates provide strong evidence of an impact of the benefit offset on the number of months in 2014 with SSDI payments. The effect is small in magnitude—an increase of 0.1 months over the 12

months in the year, accounting for only a 1 percent increase. In its direction, this finding is consistent with theory, which predicts that some treatment subjects whose earnings would be above BYA under current law will keep their earnings in that range and receive partial benefits under the offset. These individuals would have had their benefits suspended under current law. The point estimate for 2014 is highly similar to the corresponding estimate for 2013 (0.09 percentage points), which was also strongly significant (Wittenburg et al 2015, Exhibit 3-1).

As in early years, there is no evidence of impacts on total SSI benefits paid for 2014 or on the number of months with SSI payments.

9.3. Impact Variation by Beneficiary Background Characteristics

For various reasons, the benefit offset may affect outcomes for certain subgroups of SSDI beneficiaries differently than for others. To explore such a possibility, we compare earnings and benefit receipt impacts between various subgroups of Stage 1 subjects defined by beneficiary background characteristics. Each of our subgroup analyses separates demonstration participants into one of two categories based on a given background characteristic, thereby permitting us to estimate impacts for both participant categories and test whether the two impacts differ. We conduct such analyses for all nine earnings and benefit outcomes presented in Exhibits 9-1 and Exhibit 9-4. Below, we list the seven pairs of beneficiary subgroups examined.[98]

• Short-duration beneficiaries (those receiving benefits for 36 or fewer months when entering BOND)	versus	• Longer-duration beneficiaries (those receiving benefits for 37 or more months when entering BOND)
• Concurrent beneficiaries (those receiving both SSI and SSDI benefits when entering BOND)	versus	• SSDI-only beneficiaries
• Beneficiaries employed in 2010	versus	• Beneficiaries not employed in 2010
• Beneficiaries with access to Medicaid buy-in programs	versus	• Beneficiaries without access to Medicaid buy-in programs
• Younger beneficiaries (under age 50 when entering BOND)	versus	• Older beneficiaries (age 50 and older)
• Beneficiaries with a primary impairment of major affective disorder	versus	• Beneficiaries with all other primary impairments
• Beneficiaries with a primary impairment of back disorder	versus	• Beneficiaries with all other primary impairments

The subgroup analysis is considered exploratory. Hence, as with the other exploratory results presented above, any statistically significant findings are suggestive of further effects of the benefit offset. Even if the offset had no impact on any of the measures examined here, we would expect some of the impact estimates to be statistically significant solely by chance given that we conduct many hypothesis tests in this section and do not correct impact estimate *p*-values for multiple comparisons.

[98] The subgroups examined in this section were chosen before the demonstration began. Section 6.3.3 of the *Evaluation Analysis Plan* describes the motivation for examining each subgroup. Section 2.2.3 of the current report presents some predictions about how impacts might differ between subgroups.

We consider whether impacts differ according to a given background characteristic rather than whether non-zero impacts occurred within any particular subgroup defined by the background characteristic. If the impacts between paired 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. We adopted this practice (following Bloom and Michalopoulos 2013) because—lacking clear evidence that impacts differ between subgroups—the full sample yields more statistically precise findings (i.e., impact estimates with smaller standard errors) than do individual subgroups. Our ability to detect differences in the size of impacts across each pair of subgroups is necessarily limited by the sizes of the subgroup samples.

We present the full set of subgroup impact estimates in Appendix D and summarize the most important ones here. The appendix includes estimated impact differences for nine outcomes for each of seven subgroup pairs. Hence, there is a large number of tests (63) that could potentially produce multiple statistically significant findings that are spurious—i.e., that arise simply by chance in the absence of true differences in impacts between subgroups. Out of the 63 tests performed, our analysis yielded nine statistically significant differences in impact magnitude between subgroups. If all of the hypothesis tests were independent and there were in fact no true differences, we would have a 17.5 percent chance of identifying nine or more spuriously significant subgroup findings at the $p < 0.10$ level.[99] In fact, the hypothesis tests are not independent; therefore, the chance of finding nine or more spuriously significant findings in the absence of any true differences is greater—by some unknown amount—than 17.5 percent.

Of the nine statistically significant differences in impacts detected, five are significant at the $p < .10$ level, two at the $p < .05$ level, and two at the $p < .01$ level. Given the relatively small number of significant differences, it is certainly possible that some represent false signals. Those that are consistent with predictions made in the *Evaluation Analysis Plan*, as discussed below, are harder to dismiss as false signals. From Appendix D, we characterize the findings as follows:

- The benefit offset's impact on one of the earnings-related outcomes differed significantly by duration of earlier SSDI receipt (Exhibit D-1). The offset reduced the share of beneficiaries earning more than two times BYA by more for short-duration SSDI beneficiaries than for longer-duration SSDI beneficiaries. The difference in impact between the two subgroups is significant at $p < .10$.
- SSDI benefits paid differed by benefit duration status (Exhibit D-1). The offset increased total SSDI benefits paid more for long duration beneficiaries than for short duration beneficiaries. The difference in impacts is significant at $p < .10$.
- We observe two differences in earnings-related impacts by baseline employment status (Exhibit D-3). The offset reduced the share of beneficiaries earning more than two times BYA and more than three times BYA by more for those employed at baseline than for beneficiaries not employed in 2010 (Exhibit D-3). The difference in impacts is significant at $p < .10$ for two times BYA and at $p < .05$ for three times BYA.

[99] This calculation is $\sum_{k=9}^{63} \binom{63}{k} (0.1)^k (0.9)^{63-k} = 0.175$.

- Two SSDI benefit-related impacts differed by baseline employment status (Exhibit D-3). The offset increased total SSDI benefits paid and months receiving SSDI benefits in 2014 by more for beneficiaries employed in 2010 than for beneficiaries not employed that year. The difference in both impacts is significant at $p < .01$.
- Two SSDI benefit-related impacts differ by beneficiary age (Exhibit D-5). Estimated impacts on total SSDI benefits paid and number of months with SSDI payments are more positive for SSDI beneficiaries younger than age 50 at baseline than for older SSDI beneficiaries. The difference in impact is significant at $p < .10$ for total SSDI benefits paid and at $p < .05$ for number of months with SSDI payments.
- The impact on the number of months during 2014 with SSI payments is more positive for those with a back disorder as their primary impairment than for those with other primary impairments (Exhibit D-7). The difference in impact is significant at $p < .10$.

Based on results in other studies, the *Evaluation Analysis Plan* made predictions about the direction of differential subgroup impacts. Some results presented in this section are consistent with the above findings; others are not. We expected that impacts on the earnings of short-duration SSDI beneficiaries would be larger than impacts on long-duration beneficiaries because earlier research shows that beneficiaries who have recently entered SSDI are more likely to work. However, the data did not support the prediction.

We also predicted that beneficiaries employed in 2010 would be more likely to take advantage of the benefit offset than those not employed at that time. We did find, consistent with our prediction, that the impact of the offset on SSDI benefits paid and months with SSDI payments was larger for beneficiaries employed at baseline. In addition, measured impacts on the percentage of beneficiaries with earnings greater than two or three times BYA were more negative for beneficiaries employed in 2010 than for beneficiaries not employed at baseline.

Despite a strong reason to expect impacts to be smaller for concurrent beneficiaries than for SSDI-only beneficiaries, we find no statistically significant differences in impacts according to concurrent status. It is important to note, however, that analyses looking for statistically significant differences in impact between concurrent and SSDI-only beneficiaries may be able to detect only relatively large differential effects because only about 17 percent of the Stage 1 analysis sample falls into the former group (Wittenburg et al. 2012, Exhibit 3-1).

Appendix D provides evidence of impacts on individual subgroups in their own right. Some of these estimates are statistically significant at the $p < 0.10$ level, mostly echoing impacts in certain subpopulations found to be statistically significant for the entire T1 group.[100] Given the reasons cited above for expecting variation in impacts across subgroups, and given our large sample sizes, it is plausible that there will be impacts in subgroups.[101] In addition, given the large number of subgroup-

[100] Of the 126 subgroup-specific impacts in Appendix D, 42 (33 percent) are statistically significant at the $p < .10$ level. Of these 42, 40 (or 95 percent) are impacts that are observed in the full Stage 1 analysis sample results.

[101] Given the smaller sample sizes in subgroups, it is possible—and not uncommon—to have statistically significant pooled impacts and no statistically significant impact in any subgroup.

specific impacts that we examined, it is likely that some subgroup-specific impact estimates will be statistically significant even when the corresponding full-sample estimates are not significant. We are unable to determine whether those subgroup-specific estimates represent noteworthy *new* information beyond what we learned when examining the sample as a whole or whether the estimates are simply the result of chance; hence, we do not highlight those estimates here.

9.4. Summary

Overall, the impact estimates provide strong evidence that the BOND benefit offset increased total SSDI benefits paid in 2014 but no evidence that the offset affected average earnings in 2014. Exploratory findings provide evidence of positive impacts on the share of beneficiaries with employment that year and on the proportion of beneficiaries earning more than BYA. They also show a negative impact on the number of beneficiaries earning more than three times BYA. The simultaneous increase in the percentage with earnings above BYA and decline in the percentage with earnings more than three times BYA may help explain the confirmatory finding of no statistically significant impact on earnings averaged across the entire sample. Consistent with findings in the analyses of earlier years, the exploratory analysis of benefits also found a positive impact on the number of months in 2014 with SSDI payments.

We found a few differences in impacts across subgroups. Consistent with theoretical predictions, the benefit offset led to a larger reduction in the shares of beneficiaries earning above two and three times BYA among beneficiaries employed at baseline than among beneficiaries not employed at baseline. We do find that the impact of the offset on SSDI benefits paid and months with SSDI payments was greater for beneficiaries employed at baseline. Still, a much higher proportion of beneficiaries employed in 2010 had earnings two or three times BYA than beneficiaries not employed in 2010. Correspondingly, the offset led to greater increases in SSDI benefit dollars received and number of months of SSDI receipt during 2014 for beneficiaries employed at baseline versus beneficiaries not employed at baseline. For the remaining set of subgroup comparisons, we do not find statistically significant findings that enable us to confirm theoretical predictions.

10. Impacts on Outcomes Measured from Survey Data

This chapter presents impact results for all outcomes derived from the Stage 1 36-Month Survey that were not presented in earlier chapters. As in Chapter 9, this chapter presents impact estimates for T1 beneficiaries relative to C1 beneficiaries; T1 subjects' benefits are determined by the benefit offset rules while C1 subjects' benefits are determined according to current law. Except for the survey analogues of the employment, earnings, and benefit receipt outcomes examined with administrative data in Chapter 9, we have not made any theoretical predictions about how the offset will affect the survey outcomes. All statistical tests in this chapter are exploratory.[102]

Relatively few of the outcomes considered in the chapter have statistically significant impacts, that is, treatment/control differences. Specifically, of the 165 outcomes analyzed, 6 were statistically significant at the 10 percent level. Thus, a main conclusion is that impacts on outcomes measured in the survey data are either not present or are too small to be detected in the survey data. Exhibits E-8 through Exhibit E-21 in Appendix E report impact estimates for all survey outcomes discussed in this chapter.

10.1. Overview of Survey Outcome Variables

To discuss the many survey outcomes highlighted in the chapter, we organize the outcomes into 23 domains. Within each domain, the outcomes share a common focus. For instance, the work-related expenses domain contains six outcomes describing the average amount T1 and C1 beneficiaries spend on employment-related expenses: commuting, job-specific (e.g., uniforms and licenses), child care, special equipment to accommodate a disability, personal assistance services related to a disability, and total employment-related expenses.

For each outcome, we report an average for each experimental group, the difference in outcomes between the groups (point estimates of the impact), the difference's standard error, and whether the difference is statistically significant. We use Student's t-test, adjusted for the sample design, to determine whether each difference is statistically significant.

10.2. Survey Outcome Results

Our estimates of the benefit offset's effects are statistically significant for 6 of the 165 survey outcomes. Exhibits E-23 through E-25 summarize the survey impact results by outcome domain. The outcome domains with a statistically significant impact include type of business or industry, occupation, type of special equipment used, work-related expenses, specific employer accommodations, and material hardship. Statistically significant impacts are spread out across domains—no domain has more than one statistically significant impact. In most domains, we find no statistically significant impacts.

In addition to varying by outcome domain, the four statistically significant survey outcome impacts vary in direction and significance level. The four significant findings follow:

[102] That is, they are not adjusted for multiple comparisons and carry a high risk of "false positives." Hence, they may not be considered as providing confirmatory evidence about the benefit offset even when shown as statistically significant; rather, they suggest additional possible offset impacts beyond those that emerged from the exploratory analyses in Chapter 9.

- The estimated impact of the offset on working at an information business or in the information industry is negative ($p < 0.05$).
- The estimated impact of the offset on working in a community and social services occupation is positive ($p < 0.10$).
- The estimated impact of the offset on receiving other types of equipment to help employed subjects work at their current paid job is positive ($p < 0.10$).
- The estimated impact of the offset on beneficiary expenses for work-related child care is positive ($p < 0.10$).
- The estimated impact of the offset on having an employer provide modified computer software is positive ($p < 0.10$).
- The estimated impact of the offset on the share of beneficiaries who could not pay the full amount of their utility bills in the past year is negative ($p < 0.10$).

The limited number and varying domains of these statistically significant survey impacts, combined with the fact that we did not adjust the test statistics to account for the large number of hypothesis tests (165), may mean that the "significant" estimates may simply reflect chance differences between the two samples. If all of the hypothesis tests were independent and there were no true survey impacts, we would have an almost 100 percent chance of estimating four or more statistically significant impacts at the $p < 0.10$ level.[103] If we had accounted for the fact that the hypothesis tests are not independent, there would be an even higher chance of estimating four or more statistically significant impacts.

The lack of detectable impacts for the vast majority of survey outcomes does not necessarily point to an absence of impacts of sufficient size to be of interest to policymakers. Rather, it may be that, with the available sample size, we were unable to detect substantively important impacts that did occur. To illustrate, consider the point estimate for the impact on the percentage of subjects with weekly earnings above the annual equivalent of the SGA amount. Its value is 0.4 percentage points. This value represents a 17 percent increase relative to the control group value (2.4). Hence, it might be considered substantively important by some, even though it is not statistically significant—the standard error is 0.4 percentage points and the *p*-value is above the 0.10 significance threshold (Exhibit E-10).[104]

To examine whether we could detect a significant impact for a subgroup of respondents who, in theory, were more likely to use the benefit offset, we repeated the impact analysis on employment-related survey outcomes separately for beneficiaries who did and did not have earnings in 2010 (Exhibit E-22). Consistent with expectations, for those employed in 2010, we found a larger point estimate for the percentage with weekly earnings above the BYA equivalent (2.2 percentage points, or 29 percent of the

[103] This calculation is $\sum_{k=6}^{165} \binom{165}{k} (0.1)^k (0.9)^{165-k} = 0.99938$.

[104] This problem arises when the power to detect some impacts that would be of substantive interest is low. See McCloskey and Ziliak (1996) for a discussion about the distinction between statistical significance and policy or economic significance.

C1 value of 7.6 percent); however, the standard error of the estimate increased to a level (relative to the same survey outcomes measured across the entire survey sample) such that the larger point estimate is also not statistically significant because of the relatively smaller size of the subgroup of employed in 2010.

10.3. Summary

In this chapter, we presented impacts among 165 outcomes—under 23 outcome domains—measured in the Stage 1 36-Month Survey. Six of the 165 outcomes—working at an information business or in the information industry, working in a community and social services occupation, receiving other types of equipment to help employed subjects work at their current paid job, beneficiary expenses for work-related child care, having an employer provide modified computer software, and the share of beneficiaries who could not pay the full amount of their utility bills in the past year—showed some evidence of impacts at the 10 percent significance level. The statistically significant outcomes were spread across outcome domains. The lack of a greater number of statistically significant estimates may in part reflect the limited power of the survey-based impact estimators to detect impacts of substantive importance.

11. Conclusion

This chapter summarizes and discusses the findings of the first interim report on Stage 1 of BOND. The purpose of Stage 1 is to examine how a national benefit offset would affect earnings and program outcomes for the national SSDI population. To accomplish this goal, SSA randomly assigned a nationally representative sample of SSDI beneficiaries to two groups: Stage 1 treatment group beneficiaries (T1 subjects) who have their benefits adjusted under the BOND offset, and Stage 1 control group beneficiaries (C1 subjects) who continue to have their benefits paid under current law rules. The evaluation compares the experiences of the T1 subjects to those of the C1 subjects to measure the impact of the offset on annual earnings, SSDI benefits received, and other outcomes.

The analysis in this report is guided by key research questions outlined in the *Evaluation Analysis Plan* for the process, participation, and impact analyses. This report builds on previous Stage 1 reports (Wittenburg et al. 2012; Stapleton et al. 2013; Stapleton et al. 2014; Wittenburg et al. 2015; Derr et al. 2015) to answer these questions and focus on interim findings, primarily in 2014 and 2015. Specifically, the process analysis focuses on implementation of the demonstration and synthesizes data collected from qualitative interviews with WIC and EWIC staff, the BOND Implementation Team, SSA ORDES work unit staff, and beneficiaries in 2014 and 2015. The participation analysis focuses on treatment beneficiaries' participation experiences in the intervention and relies on SSA administrative data and BTS data through 2015. Finally, the impact analysis compares outcomes for T1 subjects to those of C1 subjects to assess the extent to which the benefit offset as implemented for BOND has affected beneficiaries' lives. The impact analysis is based on administrative earnings and benefits data through December 2014 and survey data focused on 2014.We highlight the most notable findings from the report here and draw cross-cutting lessons from the evidence presented as part of the respective process, participation, and impact analyses.

11.1. WIC services are generally comparable to those available to other beneficiaries, and are used at moderately higher rates

BOND makes available to T1 subjects work incentives counseling (WIC) designed to parallel similar services available to C1 subjects through sources existing outside the demonstration—except that WIC counselors advise beneficiaries about the implications of earnings for SSDI benefits under the offset rather than under current law rules. As noted in the *Process Study Report*, although WIC was implemented as designed, WIC services differed from comparable services available to C1 subjects through the Work Incentives Planning and Assistance (WIPA) program in two modest respects. First, until December 2013, all WIC counselors had responsibility for post-entitlement work needed to facilitate benefit offset adjustments; WIPA counselors have never had a comparable responsibility. In 2014, the BOND Implementation Team assumed this responsibility from WIC agencies in seven BOND sites and from an eighth in January 2015. Second, funding for the WIPA program nationally was suspended between July 2012 and July 2013. Although other resources were available to provide SSDI benefits counseling to C1 subjects throughout that period, detailed information about how benefits affect work was presumably not as widely available to C1 subjects as before or after that period. With the reinstatement of WIPA funding in July 2013 and centralization of post-entitlement work six months later, WIC and WIPA services became comparable in eight of ten sites (excluding the two sites where post-entitlement work remained the responsibility of WIC counselors). New requirements for WIPA providers as of August 2015 did not meaningfully change this alignment overall. We have no reason to think that

the idiosyncrasies of WIC service provision from 2011 through 2013 were substantial enough to have had a material effect on the impact of the BOND intervention on beneficiary earnings and SSDI benefits.

T1 subjects initially used work incentives counseling services at a modestly higher rate than other beneficiaries—presumably including C1 subjects. Based on available data for all WIPA users in 2011 (WIPA was not funded in much of 2012 and 2013, and data are not publicly available for 2014 or 2015), 1.1 percent of current-law beneficiaries used WIPA counseling services in any of the 12 months in 2011, compared to 1.3 percent of T1 subjects who used WIC services in the first eight months of the demonstration in 2011. The larger initial uptake among T1 subjects may reflect a number of factors, including the particular characteristics or personal circumstances of T1 subjects compared to beneficiaries subject to current law, heightened awareness of the availability of work incentives counseling caused by demonstration outreach to just treatment group subjects, an increased interest in employment triggered by the incentive inherent in the BOND benefit offset, and the newness and uniqueness of the benefit offset about which T1 subjects had just been informed. Through August 2015—into the fifth year of the demonstration—nearly five percent of T1 subjects had used WIC services cumulatively. Of those who had, nearly 80 percent had used services beyond information and referral services. This latter figure is consistent with the WIPA program standard of providing intensive work incentive counseling services to 80 percent of the clients it serves. We do not have data on the types and intensity of WIPA services provided to C1 beneficiaries or to all beneficiaries subject to current-law.

11.2. T1 subjects have limited knowledge of BOND earnings rules

Because the BOND benefit offset is a major change from current-law rules, to make well-informed choices about working, treatment subjects for whom the change is salient need to understand the new rules. For many SSDI beneficiaries, work is not feasible or desired, so not everyone needs to be well-informed. But for those with the capacity and desire to work, being well-informed is crucial to the success of the BOND changes. To that end, the intent of the BOND design is to ensure that T1 subjects are as well informed about the benefit offset as they would be if the same rules were part of a permanent national program. As reported in the *Stage 1 Early Assessment Report*, it appears that T1 subjects' exposure to reliable information about the benefit offset was less than intended during the first year of the demonstration. As described in the *Process Study Report*, additional outreach efforts in 2012 and 2013 were followed by a substantial increase in the share of T1 subjects who interacted with the demonstration. However, most beneficiaries did not respond to or were not reached as a part of these efforts.

Three years into the demonstration at the point of survey administration (the Stage 1 36-Month Survey) in 2014, data estimates suggest that 36 percent of T1 subjects had heard of BOND. Moreover, just an estimated 29 percent of T1 subjects understood that SSA would reduce but not completely suspend benefits were they to earn above SGA after completion of the TWP and GP. We do not know the extent to which limited beneficiary understanding reflects lack of salience of the benefit offset to the beneficiaries for whom working is not feasible or attractive as opposed to limited exposure to information about the offset. Ninety percent of T1 survey respondents see themselves as unable to work for health reasons, supporting the argument that knowledge is low because the offset is not salient to many beneficiaries. However, relatively few of those working for pay in the year before BOND entry also demonstrated understanding of the offset rules, a group for which information on the offset rules had greater salience.

Another finding from the survey suggests that T1 understanding of the offset is substantially lower than what would be the case under a fully implemented national program. Specifically, the estimated

percentage of T1 subjects with a minimal understanding of how earnings affects their benefits under the offset, 29 percent, is about half of the estimated percentage of C1 subjects with a minimal understanding of how earnings affect their benefits under current law, 54 percent. The latter group arguably provides an indicator of the share of beneficiaries that would attain a minimal understanding of the offset if it became a permanent national program, assuming that information sources about such a program would be comparable to those in existence today.

11.3. Delays in the adjustment of benefits under the offset are often substantial

As reported in the *Stage 1 Early Assessment Report*, since the start of BOND there have been challenges in administering the processes that lead up to and support benefit adjustment under the offset. Although SSA and the Implementation Team have addressed some of these challenges, delays in benefit adjustment continued throughout 2013, as described in the *Process Study Report*.

Updating those findings in this report we find that—for many of the same reasons highlighted in previous reports—delays in benefit adjustment continued throughout 2014 and 2015. One reason is that, while beneficiaries are required to report their earnings to the demonstration, some fail to do so. This delays the start of the benefit adjustment process until SSA identifies beneficiaries with earnings, typically through a review of IRS earnings records in the following year. Second, the processing of work CDRs at SSA can be slow, delaying the determination of when eligibility for payment of benefits under the offset begins and produces a longer time frame over which the beneficiary may accrue overpayments (see Section 11.4 below). Delays in work CDR processing stem primarily from insufficient resources at the ORDES BOND work unit to process cases timely. Finally, deficiencies in BSAS, the software designed to automate the final step of benefit adjustments, caused substantial delays in automated reconciliation.

In 2013, 2014, and 2015, the median time from the first month of offset use to actual payment of benefits adjusted by the offset was 22 months. This represents a significant portion of the 60 month BOND Participation Period (BPP) available to T1 beneficiaries, reducing the amount of time during which beneficiaries will fully experience the income consequences of earnings under offset rules. The adjustment time for beneficiaries whose initial benefit adjustment followed a submitted Annual Earnings Estimate (AEE) (which indicates proactive beneficiary involvement in the benefit adjustment process), was somewhat shorter—17 months; for those who did not submit an AEE, slightly longer—23 months.

Depending on the pathway of initial benefit adjustment, delays in adjustment may represent a period during which beneficiaries do not understand how earnings affect benefits under the BOND rules. In contrast, T1 subjects who entered the offset after submitting an AEE may have anticipated the effect of the offset on their benefits well in advance of the actual adjustment. The subset of T1 subjects who entered the offset via reconciliation occurring in a later calendar year without completing an AEE—52 percent of all offset users—were not necessarily aware that the benefit offset earnings rules would apply to their SSDI benefits during their initial months or years of offset use. Prompt reconciliation could have resulted in greater awareness.

11.4. The share of T1 subjects who have used the offset has grown gradually over the first five years of the demonstration

Since the opportunity to use the offset began in May 2011, the share of T1 subjects who have ever used the offset has grown gradually. As of December 2015, SSA had used offset rules to adjust the benefits of

2.8 percent of T1 subjects. An additional 2.3 percent had a cessation date in BTS as of that same date and hence will use the offset if they work and earn above BYA during the remainder of their BPP. Given the delays in adjustments noted above, it is possible that some of these beneficiaries have already qualified for a benefit payment under the offset and that SSA will retroactively adjust their benefits. Indeed, SSA first adjusted the benefits of more than half of offset users known to-date following reconciliation after the end of the calendar year.

Several beneficiary characteristics associate with adjustments under the offset in or before December 2015. Holding other characteristics constant, beneficiaries who were younger, female, had certain specific primary conditions such as neoplasms, or had either high baseline monthly SSDI benefits or high baseline average indexed monthly earnings were all more likely to have a benefit adjustment. Disabled adult children and beneficiaries with representative payees were less likely to have a benefit adjustment under the offset, all other things equal.

As of December 2015, most offset users had used the offset for more than one year. To date, over one quarter (28 percent) of offset users used the offset in just one calendar year. Almost half (45 percent) had used the offset in three or more calendar years—a duration that exceeds the median 22 month delay in benefit adjustment. Hence, the majority of these beneficiaries were presumably aware of at least the basics of how earnings affect benefits during part of the period during which they continued to use the offset.

11.5. During the first three years of BOND, SSDI benefit overpayments were more prevalent for T1 subjects than for C1 subjects, but the dollar amount of overpayments was substantially lower

Administrative data through October 2015 reveal that 83 percent of T1 offset users in 2011, 2012, and 2013 had work-related overpayments or incorrect payments (which we refer to collectively as "overpayments") in those years. Overpayments occur for reasons related to the timeliness and accuracy of benefit adjustment, including: beneficiary failure to report earnings timely, revised AEEs, inaccurate AEEs, delays in SSA processing of work CDRs, and BSAS errors.

T1 subjects were 12 percent more likely to encounter overpayments than C1 subjects during the 2011-2013 period, a statistically significant difference. This represents a 0.23 percentage point increase in prevalence over the rate of overpayments among C1 subjects, 1.88 percent. The higher prevalence of overpayments among T1 subjects may reflect several causes. First, evidence suggest that T1 subjects encountered longer delays in benefit adjustments than C1 subjects. Although we do not have comparable duration statistics for the current-law adjustment process, there is evidence that durations for C1 subjects were shorter, leaving less time for overpayments to accrue: in February 2015 only about 2 percent of pending work CDRs for current law beneficiaries had been waiting for longer than 270 days (nine months), whereas over 56 percent of treatment subjects' had work CDRs of the same age. Second, T1 beneficiaries are subject to an annual (calendar year) rather than monthly accounting period. As a result, T1 subjects may accrue overpayments due to aspects of the annual accounting period, such as inaccurate or adjusted AEEs, that do not apply to C1 subjects. Finally, it may be that a larger percentage of T1 subjects than C1 subjects had earnings above BYA or SGA, respectively, so as to necessitate downward benefit adjustments and, potentially, overpayments. That explanation—a higher percentage with sufficient earnings—would be consistent with positive point estimates for the impacts of BOND on the percentage

of T1 subjects with earnings above BYA in 2011, 2012 and 2013, although these measured differences between T1 subjects and C1 subjects were small and not statistically significant. Because of the high prevalence rate of overpayments among those with earnings above BYA (for example, 83 percent of T1 offset users in 2011, 2012, and 2013), even small differences in the percentage with earnings above this level may influence the prevalence of overpayments.

Despite their higher prevalence for T1 subjects, the average amount of overpayments to T1 subjects during the first three years of BOND was 35 percent lower than for C1 subjects—$122 versus $187 across all sample members (including those with $0 payments)—a statistically significant impact. This impact results in part from SSA overpaying by *less* for T1 subjects, who were eligible for a partial benefit payment each month compared to C1 subjects who—in the same earnings situation—were overpaid by the full amount of their benefits each month. If we assume that C1 subjects would have been overpaid at the same rate as T1 subjects if they had been assigned to the BOND treatment arm, the difference in the mean overpayment amount among those overpaid over the first three years of the demonstration was about $3,100 less for T1 subjects relative to C1 subjects. Because the stated assumption might not be exactly correct, this should be considered an upper bound for the point estimate of the mean reduction in the size of overpayments for those who would have an overpayment under the benefit offset, current law, or both.

The previous estimates apply to the first three years of the demonstration, 2011 to 2013. According to WIC and EWIC counselors, overpayments also appear to have been prevalent among treatment group subjects who first entered the offset in 2014 and 2015. However, these staff observed that overpayments during subsequent years of offset adjustment were less frequent and smaller in size. We will use administrative data to estimate overpayment rates and amounts in 2014 and 2015 in future reports.

In-depth interviews in 2015 with a sample of 20 T1 offset users revealed that 15 had experienced overpayments, although two of these overpayments pre-dated BOND. These 15 interview participants reported a range of reactions to overpayments. Some beneficiaries apparently coped well with overpayments, but others reported negative reactions or negative financial consequences, or both. Some interviewees felt they were being punished for work and some viewed overpayments as a negative feature of the benefit offset. Two interviewees attributed substantial financial distress to their overpayments once SSA began to recoup the overpayments in the form of withheld benefits. One of these two interviewees reported reducing her earnings because of the overpayments, presumably in order to increase her SSDI benefit payments. Another overpaid beneficiary reported increasing his earnings following an overpayment. The nature of the interviewee sample is such that we cannot produce estimates of the percentages of beneficiaries with various reactions to overpayments. It is apparent, though, that negative reactions of one sort or another are fairly common.

11.6. There is no detectable impact of BOND on total earnings of T1 subjects in 2014, and there is evidence of a positive impact on SSDI benefits paid in 2014

We find no evidence of a confirmatory effect of the BOND offset on total earnings of T1 subjects in 2014, the most recent year for which earnings information is available. We did, however, find exploratory evidence of impacts on the percentage employed and the percentages with earnings above two thresholds. Consistent with theory, we find statistically significant positive impacts on the percentage of T1 subjects

employed and on the percentage with earnings above BYA—coupled with a significant negative impact on the percentage of T1 subjects earning above three times BYA. Thus, although it appears that some beneficiaries changed their earnings in response to BOND, the changes are in opposite directions and may offset one another to the extent that we do not detect an impact in either direction on mean annual earnings for all beneficiaries—the study's confirmatory test for a favorable earnings effect.

Similar to previously reported findings for 2011 through 2013, we find statistically strong evidence of a positive effect of the BOND offset on total SSDI benefits *paid in* 2014, the other confirmatory outcome for the evaluation. The magnitude of the difference is about one percent higher of benefits paid to C1 subjects. Theory predicts that the offset will have (1) a negative average effect on benefits for those who would not engage in SGA under current law but who are induced to do so by the earnings incentive inherent in the BOND offset and (2) a positive average effect on benefits for those who would engage in SGA under current law. The overall positive impact of BOND for benefits paid implies that the latter effect dominates: benefit gains exceed benefit reductions on net. We will produce estimates of impacts on SSDI benefits *paid for* 2014, including retroactive adjustments that occur after the period, in the *Final Report.* Those estimates on benefits *paid for* the period will differ from impacts on SSDI benefits p*aid in* the period, perhaps substantially. However, the impact on benefits *paid for* this period are very unlikely to be negative because the measured impact on the percentage with earnings above BYA is small.

An exploratory analysis of survey data suggests that the BOND offset had little or no effect in 23 other domains of beneficiary outcomes including several health status and employment domains. In some instances, the lack of a significant finding may reflect the survey's limited power to detect potentially substantive impacts, rather than the absence of an effect. In other instances, the results may simply mean that the offset had a very limited, or no impact, on outcomes not explicitly targeted by the BOND intervention.

11.7. Taking Stock

What do we learn from the process, participation, and impact evidence presented in this report? Through 2015, WIC services appear to have been implemented generally in accord with the study's design. By the end of 2015, almost 5 percent of T1 subjects had received WIC counseling. Although T1 rates of benefit counseling appear to approximate or exceed the rates of WIPA counseling used by beneficiaries subject to current law, the percentage of T1 subjects with a minimal understanding of how earnings affects their benefits under the offset, estimated at 29 percent, is about half of the estimated percentage of C1 subjects with a minimal understanding of how earnings affect their benefits under current law, 54 percent.

SSA had reduced benefits under the offset in at least one month for 2.8 percent of T1 subjects. Throughout the demonstration, the number of offset reductions has lagged behind the number of offset users for whom SSA should have—and eventually will—adjust benefits under the offset. In 2013, 2014, and 2015, the median time from the first month in which a beneficiary's earnings first met the criteria for offset use to an actual benefit offset adjustment was 22 months. This delay has two important consequences. First, it reduces the time period over which beneficiaries fully experience the income consequences of earnings under offset rules and hence may contribute to beneficiaries' incomplete understanding of how earnings affect benefits under the BOND benefit offset. Second, beneficiaries who continue to earn above SGA during the lag period will accrue overpayments. Indeed, 83 percent of T1 offset users in 2011, 2012, and 2013 had an overpayment.

The impact analysis provides no evidence of an effect of the offset on the evaluation's first confirmatory outcome, 2014 annual earnings. Exploratory evidence suggests underlying earnings changes in opposite directions: an increase in the percentage of T1 beneficiaries employed and earning above BYA coupled with a decrease in the percentage with earnings below three times BYA. Both of these shifts involve less than one percent of all T1 beneficiaries. The next Stage 1 snapshot report will examine how the benefit offset affects earnings, employment, and benefits in 2015. In a final report, we will evaluate the same set of outcomes through 2016.

References

Bell, Stephen H., Daniel Gubits, David Stapleton, David Wittenburg, Michelle Derr, Arkadipta Ghosh, and Sara Ansell. "BOND Implementation and Evaluation. Evaluation Analysis Plan." Final Report Submitted to Social Security Administration. Cambridge, MA: Abt Associates, March 2011.

Bloom, Howard S., Carolyn J. Hill, and James A. Riccio. "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments." *Journal of Policy Analysis and Management*, vol. 22, no. 4, 2003, pp. 551-575.

Burkhauser, Richard V., Andrew J. Houtenville, and David C. Wittenburg. "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, pp. 23-86. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2003.

Card, David, Jochen Kluve, and Andrea Weber. "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations." Working Paper 21431. Cambridge, MA: National Bureau of Economic Research, 2015.

Chromy, James R. "Sequential Sample Selection Methods." *Proceedings of the Section on Survey Research Methods of the American Statistical Association*. Alexandria, VA: American Statistical Association, 1979.

Derr, Michelle, Denise Hoffman, Jillian Berk, Ann Person, David Stapleton, Sarah Croake, Christopher Jones, and Jonathan McCay. "BOND Implementation and Evaluation: Process Study Report." Submitted to the Social Security Administration, Office of Research, Demonstrations, and Employment Support. Cambridge, MA: Abt Associates, February 2015.

Greenberg, David H., Charles Michalopoulos, and Philip K. Robins. "A Meta-Analysis of Government-Sponsored Training Programs." *Industrial and Labor Relations Review*, vol. 57, no. 1, 2003, pp. 31-53.

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

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

Gubits, Daniel, Judy Geyer, Denise Hoffman, Sarah Croake, Utsav Kattel, David Judkins, Stephen Bell, and David Stapleton. "BOND Implementation and Evaluation: 2015 Stage 2 Interim Process,

Participation, and Impact Report." Submitted to the Social Security Administration. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, 2016.

Honeycutt, Todd, and David Stapleton. "Striking While the Iron Is Hot: The Effect of Vocational Rehabilitation Service Wait Times on Employment Outcomes for Applicants Receiving Social Security Benefits." *Journal of Vocational Rehabilitation*, vol. 39, no. 2, 2013, pp. 137-152.

Houtenville, Andrew J., Debra L. Brucker, and Eric A. Lauer. *Annual Compendium of Disability Statistics: 2015*. Durham, NH: University of New Hampshire, Institute on Disability, 2016.

Houtenville, Andrew J., and Tony Ruiz. *Annual Compendium of Disability Statistics: 2012*. Durham, NH: University of New Hampshire, Institute on Disability, 2012.

Johnson, Nicholas, Phil Oliff, and Erica Williams. "An Update on State Budget Cuts." Washington, DC: Center on Budget and Policy Priorities, 2011.

Liu, Su, and David C. Stapleton. "Longitudinal Statistics on Work Activity and Use of Employment Supports for New Social Security Disability Insurance Beneficiaries." *Social Security Bulletin*, vol. 71, no. 3, 2011, pp. 35–60.

Livermore, Gina A., Denise Hoffman, and Maura Bardos. "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, 2012.

McCloskey, Deirdre N., and Stephen T. Ziliak. "The Standard Error of Regressions." *Journal of Economic Literature*, vol. 34, no. 1, 1996, pp. 97-114.

Rehabilitation Research and Training Center on Disability Statistics and Demographics. *Annual Compendium of Disability Statistics: 2009*. New York, NY: Hunter College, 2009.

Schimmel, Jody, Sarah Prenovitz, Gina Livermore, and Alex Bryce. "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, 2013.

Schochet, Peter. "The Estimation of Average Treatment Effects for Clustered RCTs of Education Interventions." Washington, DC: U.S. Department of Education, Institute of Education Sciences, March 2009.

Social Security Administration. *Annual Statistical Report on the Social Security Disability Insurance Program, 2011*. Baltimore, MD: SSA, 2012.

Social Security Administration. *Monthly Statistical Snapshot, December 2012*. Baltimore, MD: SSA, 2013a.

Social Security Administration. *OASDI and SSI Program Rates & Limits, 2014.* Washington, DC: SSA, 2013b.

Social Security Administration Office of the Inspector General. *Overpayments in the Social Security Administration's Disability Programs—A 10-Year Study.* Baltimore, MD: SSA, 2015.

Social Security Administration. "Program Operations Manual System (POMS): GN 02201.013. Administrative Tolerances Title II Overpayments." Available at https://secure.ssa.gov/poms.nsf/lnx/0202201013. Accessed February 29, 2016.

Social Security Administration. *Annual Statistical Supplement to the Social Security Bulletin, 2015*. Washington, DC: SSA, 2016.

Stapleton, David, Stephen Bell, David Wittenburg, Brian Sokol, and Debi McInnis. "BOND Final Design Report." Submitted to the Social Security Administration. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, 2010.

Stapleton, David, David Wittenburg, Daniel Gubits, David Judkins, David Mann, and Andrew McGuirk. "BOND Implementation and Evaluation: First-Year Snapshot of Earnings and Benefit Impacts for Stage 1." Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, 2013.

Stapleton, David, David Wittenburg, David R. Mann, Denise Hoffman, and Andrew McGuirk. "BOND Implementation and Evaluation: Second-Year Snapshot of Earnings and Benefit Impacts for Stage 1." Submitted to the Social Security Administration, Office of Program Development & Research. Washington, DC: Mathematica Policy Research, 2014.

Weathers, Robert and Jeffrey Hemmeter. "The Impact of Changing Financial Work Incentives on the Earnings of Social Security Disability Insurance (SSDI) Beneficiaries." *Journal of Policy Analysis and Management*, vol. 30, no. 4, 2011, pp. 708-728.

Williams, R.L., and J.R. Chromy. "SAS Sample Selection MACROS." *Proceedings of the Fifth International SAS Users Group International Conference*. Cary, N.C.: SAS Institute, 1980.

Wittenburg, David, David Stapleton, Michelle Derr, Denise W. Hoffman, and David R. Mann. "BOND Implementation and Evaluation: BOND Stage 1 Early Assessment Report." Submitted to the Social Security Administration, Office of Program Development & Research. Cambridge, MA: Abt Associates, May 2012.

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

Appendix A. Impact Methodology for Administrative Outcomes

This report presents impact estimates for outcomes derived from (i) administrative data sources and (ii) the Stage 1 36-Month Survey. This appendix (Appendix A) provides details about the impact methodology for administrative outcomes; the next appendix (Appendix B) provides details about the impact methodology for outcomes derived from the Stage 1 36-Month Survey.

We made two small changes to the impact estimation methodology for administrative outcomes relative to that used to estimate impacts on administrative outcomes in the *Third-Year Snapshot of Earnings and Benefits Impacts for Stage 1*.[105] The two changes are (1) a correction to definitions of two covariates used in the estimation model and (2) a correction to the analysis weights. This appendix provides details of these two changes. This appendix also describes in detail the measure of earnings used to generate earnings-related administrative outcomes (A.3). In A.4, we describe a key statistical test in the subgroup analysis (inadvertently omitted from previous Stage 1 reports). The final section of the appendix (A.5) presents revised impact results with the new covariates and analysis weights. The resulting changes to the earlier estimates are trivial and do not in any way affect the substantive findings.

A.1 Change in two model covariates

The estimation procedure, multiple comparisons adjustment, and full set of covariates for impacts on administrative outcomes were described in Appendix A of the *Third-Year Snapshot of Earnings and Benefits Impacts for Stage 1*. During quality review of results for the current report, we discovered that two covariates—indicators for "ineligible for Stage 2 for geographical reasons" and "ineligible for Stage 2 for having a legal guardian who was not a representative payee"—were defined based on a data extract that included information for C1-supplement subjects but not for T1 and C1-core subjects. These two covariates have been re-created based on a data extract with information for the entire Stage 1 sample.

A.2 Change in analysis weights

During quality review of the survey weights (see Appendix B), we discovered a small imbalance between the T1 and C1 groups when weighted with the administrative analysis weights.[106] Upon further investigation, we learned that the construction of the original set of administrative analysis weights (described in the *First-Year Snapshot of Earnings and Benefits Impacts for Stage 1*) did not fully capture the process by which the Implementation Team randomly selected subjects for the Stage 2 outreach waves. This had implications for the Stage 1 administrative analysis weights because subjects originally assigned to the Stage 2 recruitment pool who were not included in the random replicates that were released for Stage 2 outreach waves were placed in the C1 group. These discoveries led to the construction of a revised set of administrative analytic weights. Most of the changes in the construction of

[105] Specifically, the analysis sample, the estimation procedure, the multiple comparisons adjustment, and the subgroup analysis methods for administrative outcomes are the same as those used for the administrative outcomes in the *Third-Year Snapshot of Earnings and Benefits Impacts for Stage 1*.

[106] Although a test using the original set of weights (presented in Exhibit 2-2 of the *First-Year Snapshot of Earnings and Benefits Impacts for Stage 1*) showed balance between the T1 and C1 groups, that test only used a limited set of key covariates. Recent testing of other covariates revealed some imbalance between the groups.

the weights are made to better recover the true probability of selection to the Stage 2 outreach, and therefore to better recover the probability of being selected to the C1 group. Compared to the original method of weight construction, the new method:

- Accounts for the fact that probability of selection for Stage 2 outreach varied by SSDI entitlement month for subjects whose SSDI entitlement started between January 2008 and July 2009 (the most recently entitled subjects had were more likely to be selected for outreach);[107]
- Accounts for the possibility of Stage 2 eligibility status changing between December 2010 and April 2011 sample files;
- Accounts for some Stage 2 outreach waves being limited to certain geographical zones within sites;
- Revises how probability of assignment to the C1 group is calculated;
- Removes concurrent status as a strata-defining characteristic for selection to T1 and the C1-core;[108] and
- Adds post-stratification steps to correct for chance imbalances in weights caused by the contamination adjustment.

Collectively, these changes in the construction of weights resulted in a revised set of administrative weights that is largely similar to the original set of weights. The correlation between the original and revised weights is .997.

The full details of analysis weight construction are as follows.

The analysis weights reflect the probabilities of selection into the T1 and C1 groups. Initially, 10 SSA area offices were randomly selected as BOND sites. Next, BOND-eligible subjects were randomly selected for T1 and the C1-core. (The C1-core is a group of Stage 1 control subjects that is the same size as the T1 group—about 80,000 subjects—and selected in an identical manner.) Subjects who were not selected into T1 or the C1-core and who *were not* eligible for Stage 2 outreach (about 200,000 subjects) were placed in the C1 group as part of the C1-supplement. Of the roughly 800,000 subjects who were not selected into T1 or the C1-core and who *were* eligible for Stage 2 outreach, about 200,000 were randomly selected into Stage 2 outreach waves and the remaining 600,000 were placed in the C1 group, also as part of the C1-supplement. About 9,000 subjects were removed from Stage 1 as part of the "contamination" adjustment.[109]

[107] The variation in selection probabilities by entitlement month was caused by the short/long entitlement duration status used for oversampling being determined as of the time of each Stage 2 outreach wave (i.e., the dividing line for short/long status was different for each outreach wave).

[108] Concurrent status was removed as a strata-defining characteristic because once Stage 2 eligibility is taken into account, selection probabilities did not depend on concurrent status. Excluding concurrent status from the construction of weights removes unnecessary variability in the analysis weights.

[109] The "contamination" adjustment is described in the *First-Year Snapshot of Earnings and Benefits Impacts for Stage 1*, Appendix A, Section A.4.

The first component of the analysis weight is the reciprocal of the probability of site selection. As explained in Stapleton et al. (2010), 10 SSA area offices were selected as sites for BOND from eight strata defined by census region (Northeast, Midwest, South, or West) and proportion of beneficiaries living in Medicaid buy-in states (low or high). A single area office was selected from each stratum, with one exception; three area offices were selected from the low Medicaid Buy-in stratum in the South region, which had many more area offices and beneficiaries than the other strata. [110] The area offices were selected in each stratum using probability proportional to size systematic sampling, in which size is defined as the number of SSDI beneficiaries served by the area office.

The second component of the analysis weights is the reciprocal of the probability of selection into T1 or C1 assignment groups conditional on selection of site. The weights are constructed differently for T1 and C1 subjects, reflecting the different processes for being randomly selected into T1 or C1. Subjects could only be selected into the T1 group during the initial random selection of T1 and C1-core. Subjects could be selected into C1 either through selection into the C1-core or through selection into the C1-supplement.

The random selection to the T1 or C1-core occurred within 20 strata (within BOND sites) defined by site and the distinction between short-duration entitlement to SSDI benefits (36 months or fewer as of June 1, 2011) versus longer-duration entitlement (37 months or more as of June 1, 2011). Short-duration beneficiaries were oversampled such that they made up 50 percent of the T1 group and the C1-core (rather than their naturally occurring proportion of about 30 percent).

The random selection of Stage-2 eligible subjects into each outreach wave occurred within strata defined by geographical zone and by the distinction between short- and long-duration entitlement *as of the mailing date* for the wave. (Initial outreach mailing began in January 2011 and ended in May 2012.) For logistical reasons, some outreach waves were limited to certain geographical zones within sites. Most waves oversampled short-duration subjects in order to insure that at least half of Stage 2 volunteers were short-duration subjects (a goal of the Stage 2 design). The degree of oversampling differed across waves. Stage 2 eligibility for the first three outreach waves (known as the Stage 2 "pilot") was based on the December 2010 sample file and eligibility for the remaining outreach waves was based on the subsequent April 2011 sample file. Probability for random selection to Stage 2 outreach thus differed according to three factors:

a) SSDI entitlement start month (which determined short-/long-duration status for each outreach wave),
b) Geographical zone of residence, and
c) Stage 2 eligibility status in the December 2010 and April 2011 files.

Below, we specify weights (prior to post-stratification) separately for (1) T1 subjects who are unrelated to other BOND subjects within site, (2) C1 subjects who are unrelated to other BOND subjects within site,

[110] Because three area offices were selected from this stratum, the first component of all analysis weights for sample members from this stratum is $\left(\frac{N_m}{3N_{mk}}\right)$, rather than $\left(\frac{N_m}{N_{mk}}\right)$.

and (3) Stage 1 subjects who are related to another subject in the same site and assignment group. Each Stage 1 sample member who is unrelated to other BOND subjects in their site is assigned an analysis weight given by:

T1 subjects who are unrelated to other BOND subjects within site

$$w_{mkji}^{T1} = \left(\frac{N_m}{N_{mk}}\right)\left(\frac{N_{mkj}}{N_{mkj}^{T1}}\right)$$

where:

- w_{mkji}^{T1} is the Stage 1 analysis weight for a subject *i* who is served by site *k* within national stratum *m*, is a beneficiary of type *j* (short- or long-duration as of June 1, 2011), and has been randomly assigned to T1,
- N_m denotes the number of SSDI beneficiaries in national stratum *m* (*m*= 1, …, 8),
- N_{mk} denotes the number of SSDI beneficiaries served by site *k* within national stratum *m*,
- N_{mkj} denotes the number of BOND-eligible SSDI beneficiaries served by site *k* within national stratum *m* who are of type *j*,
- N_{mkj}^{T1} denotes the number of SSDI beneficiaries of type *j* in site *k* within national stratum *m* who were randomly assigned to T1.

C1 subjects who are unrelated to other BOND subjects within site

$$w_{mkj\ell i}^{C1} = \left(\frac{N_m}{N_{mk}}\right)\left(\frac{1}{Pr(selected\ to\ C1)^{mkj\ell}}\right)$$

$$Pr(selected\ to\ C1)^{mkj\ell} = Pr(selected\ to\ C1core)^{mkj} + Pr(selected\ to\ C1supplement)^{mkj\ell}$$

$$Pr(selected\ to\ C1core)^{mkj} = \left(\frac{N_{mk}}{N_m}\right)\left(\frac{N_{mkj}^{C1core}}{N_{mkj}}\right)$$

$$Pr(selected\ to\ C1supplement)^{mkj\ell} = Pr(not\ in\ core)^{mkj} Pr(not\ in\ outreach|not\ in\ core)^{mkj\ell}$$

$$Pr(not\ in\ core)^{mkj} = \left(1 - Pr(selected\ to\ T1)^{mkj} - Pr(selected\ to\ C1core)^{mkj}\right)$$

$$Pr(not\ in\ outreach|not\ in\ core)^{mkj\ell} = 1 - Pr(in\ outreach|not\ in\ core)^{mkj\ell}$$

where:

- N_m, N_{mk}, and N_{mkj} are defined as above,

- $w_{mkj\ell i}^{C1}$ is the Stage 1 analysis weight for a subject *i* who is served by site *k* within national stratum *m*, is a beneficiary of type *j* (short- or long-duration as of June 1, 2011), and of stratum ℓ (defined by SSDI entitlement start month, geographical zone, and Stage 2 eligibility status in December 2010 and April 2011 files), and has been randomly assigned to C1,
- $Pr(selected\ to\ C1)^{mkj\ell}$ is the unconditional probability of being randomly selected to C1 for subjects in site *k* within national stratum *m* who are of type *j* and stratum ℓ,
- $Pr(selected\ to\ C1core)^{mkj}$ is the unconditional probability of being randomly selected to the C1-core for subjects in site *k* within national stratum *m* who are of type *j* (note that this is defined in the same manner as the probability of selection to T1) ,
- $Pr(selected\ to\ C1supplement)^{mkj\ell}$ is the unconditional probability of being randomly selected to the C1-supplement for subjects in site *k* within national stratum *m* who are of type *j* and stratum ℓ,
- $Pr(not\ in\ outreach|not\ in\ core)^{mkj\ell}$ is the conditional probability of *not* being randomly selected to Stage 2 outreach conditional on not being randomly selected to T1 or C1-core for subjects in site *k* within national stratum *m* who are of type *j* and stratum ℓ,
- $Pr(selected\ to\ T1)^{mkj}$ is the unconditional probability of being randomly selected to T1 for subjects in site *k* within national stratum *m* who are of type *j* –the reciprocal of the within-site component of w_{mkji}^{T1} (defined above),
- $Pr(in\ outreach|not\ in\ core)^{mkj\ell}$ is the conditional probability of being randomly selected to Stage 2 outreach conditional on not being randomly selected to T1 or C1-core for subjects in site *k* within national stratum *m* who are of type *j* and stratum ℓ.

Because of the large number of cells involved, direct estimation through sample counts of $Pr(in\ outreach|not\ in\ core)^{mkj\ell}$ would produce noisy estimates of probabilities. Therefore, we estimate these conditional probabilities with site-level logistic regression models of the form:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \boldsymbol{\beta_1 SSDIStartMonthCohort} + \boldsymbol{\beta_2 GeographicalZone} + \boldsymbol{\beta_3 Stage2Eligibility} + \boldsymbol{\beta_4}(\boldsymbol{SSDIStartMonthCohort}\ X\ \boldsymbol{GeographicalZone}) + \varepsilon$$

where:

- π is the unknown probability of being randomly selected into outreach,
- $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, $\boldsymbol{\beta_3}$, and $\boldsymbol{\beta_4}$ are vectors of coefficients,
- $\boldsymbol{SSDIStartMonthCohort}$ is a set of dummy variables indicating the SSDI entitlement start month cohort,[111]

[111] ***SSDIStartMonthCohort*** is a set of dummy variables representing 20 levels: January 2008 and before, each month from February 2008 to May 2009, May 2008 and before, June 2008 and after, and June 2009 and after. The number of pertinent levels for each subject within the ***SSDIStartMonthCohort*** set is determined by the

- ***GeographicalZone*** is a set of 3, 4, or 5 dummy variables depending on the site indicating geographical zone within site,[112]
- ***Stage2Eligibility*** is a set of 3 dummy variables: "eligible in both sample files", "eligible in December 2010 file only", and "eligible in April file only".

The value of $Pr(in\ outreach|not\ in\ core)^{mkj\ell}$ is 0.0 for subjects who were not eligible for Stage 2 in either sample file.

Stage 1 subjects who are related to another subject in the same site and assignment group

In essence, the above expressions for T1 and C1 weights are the product of a site weight and a within-site weight. Using this terminology, we can define the analysis weight of Stage 1 sample members who are related to another subject in the same assignment group as the product of the common site weight and the within site weights of each of the related sample members. In notation, this is:

$$w^{g}_{mkj\ell i} = \left(\frac{N_m}{N_{mk}}\right)\left(\frac{1}{Pr(selected\ to\ g)^{mkj\ell}_{i}}\right)\left(\frac{1}{Pr(selected\ to\ g)^{mkj\ell}_{r}}\right)$$

where:

- N_m and N_{mk} are defined as above,
- $w^{g}_{mkj\ell i}$ is the Stage 1 analysis weight for a subject *i* who is served by site *k* within national stratum *m*, is a beneficiary of type *j* (short- or long-duration as of June 1, 2011), and of stratum ℓ (defined by SSDI entitlement start month, geographical zone, and Stage 2 eligibility status in December 2010 and April 2011 files), and has been randomly assigned to group *g* (T1 or C1),
- $Pr(selected\ to\ g)^{mkj\ell}_{i}$ is the unconditional probability of being randomly selected to either T1 or C1 (defined above), with subscript *i* added to emphasize that this is the type *j* and stratum ℓ of beneficiary *i*,

number of outreach waves a subject could have been selected into based on their Stage 2 eligibility in the two sample files. Subjects with different values for the ***SSDIStartMonthCohort*** set had different probabilities of selection into outreach because their short-/long-duration status differed in the random selection of at least one outreach wave. For subjects who were eligible for Stage 2 in both sample files, the ***SSDIStartMonthCohort*** set has 18 pertinent levels: January 2008 and before, each month from February 2008 to May 2009, and June 2009 and after. For subjects who were eligible for Stage 2 only in the December 2010 sample file, the ***SSDIStartMonthCohort*** set has 6 pertinent levels: January 2008 and before, each month from February 2008 to May 2008, and June 2008 and after. For subjects who were eligible for Stage 2 only in the April 2011 sample file, the ***SSDIStartMonthCohort*** set has 14 pertinent levels: May 2008 and before, each month from June 2008 to May 2009, and June 2009 and after.

[112] The Greater Detroit and Greater Houston sites have 3 zones, the DC Metro site has 4 zones, and all other sites have 5 zones.

- $Pr(selected\ to\ g)_r^{mkj\ell}$ is the unconditional probability of being randomly selected to either T1 or C1 (defined above) for subjects with the type *j* and stratum ℓ of beneficiary *r*, who is the related family member of beneficiary *i*.

Note that related family members (beneficiary *i* and beneficiary *r*) who remain in the sample always are from the same stratum *m*, site *k*, and group *g* (otherwise they have been removed from the analysis sample). The related family members may differ only according to type *j* and stratum ℓ.

Post-stratification adjustments

The final steps of the construction of the analysis weights are two post-stratification adjustments to correct for chance imbalances in weights caused by the contamination adjustment (that is, caused by the exclusion of related subjects where the number of related subjects within the family unit is three or greater, the exclusion of pairs of related subjects where subjects are in different assignment groups ["contaminated"], and the special weighting needed for pairs of related subjects who are in the same assignment group ["uncontaminated"]).[113] The post-stratification adjustments use a universe of the final Stage 1 sample (including "uncontaminated" pairs of related subjects) plus "contaminated" pairs of related subjects. The Stage 1 weights of these subjects (including those in family pairs) are calculated as if none of the subjects were related (i.e., according to the definitions above for unrelated subjects). Then two sets of universe sums (also known as "control totals" in the statistics literature) are calculated:

1) The total weight of unrelated subjects and the total weight of subjects in family pairs, and
2) The total weight of subjects in each site.

The first post-stratification adjustment scales the post-contamination adjustment weights defined above so that, in both T1 and C1: 1) unrelated subjects have the same total weight as the unrelated subjects in the universe prior to the contamination adjustment, and 2) non-removed family pair subjects have the same total weight as all family pair subjects in the universe prior to the contamination adjustment.

The second post-stratification adjustment scales the weights after the first post-stratification adjustment so that, in both T1 and C1, subjects in a site have the same total weight as the universe subjects in that site. This insures that, in each site, the total weight of T1 subjects is equal to the total weight of C1 subjects.

A.3 Social Security Earnings

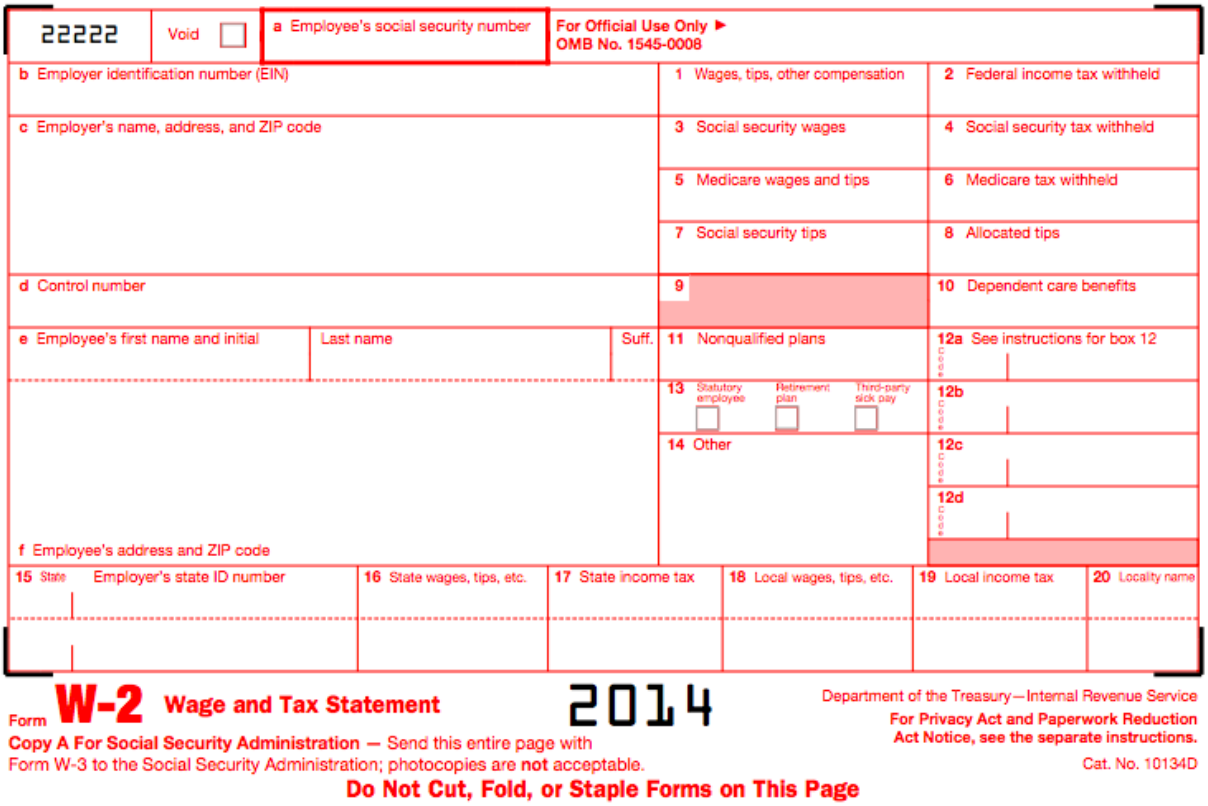
The Social Security Administration made the Summary Segment of the Master Earnings File (MEF) available to this evaluation. The MEF is SSA's primary repository of earnings data for the US population. The MEF contains all information from the W-2 forms submitted annually by employers to SSA for each paid employee and the relevant information for calculating benefits from 1040-SE forms that self-employed individuals send to the IRS. The Summary Segment summarizes a limited set of data from the MEF. Therefore a limited set of information is available to the BOND evaluation. For example, the Summary Segment does not include total earnings subject to income tax. Rather, the Summary Segment contains data on annual earnings that are subject to Old-Age, Survivors, and Disability Insurance (OASDI) taxes, otherwise known as Social Security taxes. The revenue from OASDI taxes funds

[113] Related subjects are defined as subjects in the same site who are entitled to SSDI benefits on the basis of the work history of a common primary beneficiary.

insurance benefit payments to retired workers and their spouses and children; survivors of deceased workers; and disabled workers and their spouses and dependent children. We next describe how Social Security taxes are reported to SSA.

The W-2 form lists several types of earnings amounts (Exhibit A-1). Box 1 reports an employee's total wage, tips, and other compensation from a specific employer that is subject to income tax. Several types of wages are excluded from Box 1, such as payments to retirement accounts (401Ks). Social security taxable earnings are reported on W-2 forms in Box 3 ("Social security wages") and Box 7 ("Social security tips"); payments to retirement accounts are taxed, and therefore included. Social Security taxable earnings are capped at a maximum. For 2014, the maximum was $117,000 (SSA 2013b). Amounts above this maximum are not subject to Social Security taxes, and thus the sum of Box 3 and Box 7 will never exceed the maximum, regardless of what is reported in Box 1. The sum of Box 3 and Box 7 could be less than Box 1 (for example, because wages exceed the wage base limit). However, the sum of Box 3 and Box 7 could also be more than Box 1 (for example, payments to retirement accounts and dependent care accounts are taxable for Social Security in the year they are earned).

Exhibit A-1. W-2 Wage and Tax Statement



The Summary Segment of the MEF contains the summed total of the Social Security earnings amounts from all of the W-2 forms (Box 3 and Box 7) and the 1040-SE form posted to the MEF. Therefore, the summed totals of Social Security earnings amounts are the data available to the BOND evaluation. There

are some disadvantages to relying on Social Security earnings as an overall earnings measure. Social Security earnings may be different from all employment income for the following reasons:

i. Not all jobs are covered by Social Security. Non-covered jobs include some state and local government positions and railroad workers. Only six percent of the US workforce does not participate in Social Security (SSA 2016). For example, teachers in some states do not pay Social Security taxes on their earnings. Of the BOND sites, teachers in Colorado, Maine, and Massachusetts fall into this category.

ii. For each W-2 and 1040-SE form, Social Security earnings are capped at a maximum taxable amount, $117,000 for 2014. However, we do not expect this fact to be a problem for the analysis because very few have earnings at or above that amount. In 2014, 0.03 percent of Stage 1 subjects had earnings equal to the 2014 maximum taxable amount and 0.01 percent had earnings above the 2014 maximum taxable amount. In addition, beneficiaries who are earning at or above that amount are unlikely to have a behavioral response to the offset.

iii. Not all work and earnings will be reported on a W-2 or 1040-SE form (i.e. "under-the-table" earnings).

As the earnings data available on the Summary Segment of the MEF do not include all earnings countable towards SGA, our estimates of earnings, employment, and proportion working above BYA may have a small downward bias compared to measures defined by total earnings countable towards SGA.[114] In addition, the estimate of the impact 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 (item number one in the list, above). On the other hand, the estimate could have a slight upward bias due to the fact that the offset may induce some people with under the table earnings to report them. Measures of weekly earnings and employment taken from survey data should not be subject to the same source of bias (though they are subject to other biases; in particular, recall bias and non-response bias).

A.4 Subgroup Analysis

The current report discusses subgroup analysis of impacts on administrative outcomes in Chapter 9 and presents results in Appendix D. The impacts for each subgroup are estimated separately with the same estimation procedure as that used for the full sample (described in the *Third-Year Snapshot of Earnings and Benefits Impacts for Stage 1*).[115]

[114] Not available for this evaluation, the Social Security Administration also has records of Box 1 earnings in the Detailed Segment of the MEF. Still, Box 1 earnings data would not offer a complete picture of earnings countable towards SGA because not all work and earnings are reported on a W-2 or 1040-SE form.

[115] Specifically, we estimate a working model in the full sample control group and calculate residuals for both treatment and control subjects based on the control group regression coefficients. Then, within each subgroup separately, we form random groups and regress average group residuals on treatment status.

Additionally, we tested for difference in impacts between subgroups. We estimated the variance of the difference in subgroup impacts in two ways and used the larger standard error for statistical testing.[116] In the first way, sites were treated as strata (the "conditional" analysis). In the second way, sites were treated as randomly selected clusters (the "unconditional" analysis). The p-value for the test that the difference in impacts between subgroups is equal to 0 was calculated using the estimated difference from the unconditional analysis and the larger of the two variance estimates. This was done to reduce the number of false positives created by noisy estimates of the unconditional variances. See Appendix A of the *Third-Year Snapshot of Earnings and Benefits Impacts for Stage 1* for more details on this general issue.

A.5 Revised Administrative Impact Estimates

The revised specifications and weights described in Sections A.1 and A.2 produce trivial changes to estimates relative to the earlier specifications and weights. Exhibit A-2 compares the two sets of estimates for the confirmatory estimates on total annual earnings and total SSDI benefits paid for 2011, 2012, 2013, and 2014. There are slight changes to the impact estimate of no more than $2 and corresponding changes in standard errors are also small. There are no changes in the level of statistical significance.

In the remainder of this section, we present revised estimates for administrative impacts for 2011, 2012, and 2013.[117] Stapleton et al. (2013), Stapleton et al. (2014), and Wittenburg et al. (2015), respectively, presented estimates using the earlier specifications and weights. Using the revised specification and weights, Exhibit A-2 (for 2011), Exhibit A-5 (for 2012), and Exhibit A-8 (for 2013) present revised estimates. Relative to the earlier estimates, the revised estimates are similar in magnitude and identical in level of statistical significance to the originally published estimates. There were two changes in levels of statistical significance among the 54 subgroup outcomes previously presented for 2011, 2012, and 2013:

- The difference in earnings above twice BYA for 2012 for short- and long-duration beneficiaries was not statistically significant according to the earlier specification and weights. With the revised specification and weights, the difference is statistically significant at the 10 percent level (Exhibit A-6). The impact estimate changed from -0.19 to -0.18.
- The difference in employment during 2012 for SSDI-only and concurrent beneficiaries was not statistically significant using the earlier specification and weights. The revised specification and weights, it is statistically significant at the 10 percent level (Exhibit A-7). The impact estimate changed from 0.62 to 0.67.

[116] In both analyses, the weighted average residuals from the random groups formed in the two subgroups (see description of full-sample estimation procedure in the *Third-Year Snapshot of Earnings and Benefits Impacts for Stage 1*) are regressed on an indicator for treatment status, an indicator for subgroup membership, and an interaction term between treatment status and subgroup membership. The coefficient on the interaction term is the estimate of the difference in impacts between the subgroups. The analyses use the PROC SURVEYREG procedure in the SAS software package.

[117] Corresponding administrative impacts for 2014 are available in Chapter 9 of this report.

Exhibit A-2. Comparison of Estimated Impacts on Total Earnings and Total SSDI Benefits Paid for 2011-2014 Based on Original and Revised Specification and Weights

Outcome	Original Impact Estimate	Revised Impact Estimate	Difference
Total Earnings			
Total 2011 earnings (January – December 2011)	-$9 ($25)	-$9 ($24)	$0
Total 2012 earnings (January – December 2012)	$7[a] ($31)	$8 ($29)	-$1
Total 2013 earnings (January – December 2013)	$2 ($23)	$4 ($23)	-$2
Total 2014 earnings (January – December 2014)	$18 ($28)	$17 ($27)	$0
Total SSDI Benefits Paid			
Total SSDI benefits paid in 2011 (May – December 2011)	$23# ($10)	$24# ($11)	-$1
Total SSDI benefits paid in 2012 (January – December 2012)	$69### ($14)	$70### ($17)	$0
Total SSDI benefits paid in 2013 (January – December 2013)	$139### ($23)	$138### ($22)	$1
Total SSDI benefits paid in 2014 (January – December 2014)	$132### ($29)	$132### ($28)	$0

Source: SSA administrative records for calendar years 2011-2014 and baseline SSA administrative data (used as covariates in impact analysis regression equations).

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail). Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics. Original impact estimates for 2011, 2012, and 2013 were presented in Stapleton et al. (2013), Stapleton et al. (2014), and Wittenburg et al. (2015). Revised impact estimates include a change in two model covariates and use revised administrative analysis weights, as described in Sections A.1 and A.2. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: T1 = 77,115, C2 = 891,598

[a]Due to a rounding mistake, this value was incorrectly reported as $6 in Exhibit 3-1 of Stapleton et al. (2014), rather than $7.

#/###/### 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).

Exhibit A-3. 2011 Impact Estimates on Earnings and Benefits Outcomes Based on Revised Specifications and Weights

Outcome	T1 Mean	C1 Mean	Impact Estimate
Employment (January–December 2011)			
Total earnings (confirmatory)	$1,193	$1,201	-$9 ($24)
Employment during year (%)	16.13	16.02	0.11 (0.11)
Earnings above BYA (%)	2.43	2.40	0.03 (0.11)
Earnings above 2x BYA (%)	0.95	0.98	-0.04 (0.05)
Earnings above 3x BYA (%)	0.53	0.54	0.00 (0.03)
Benefit Receipt (May–December 2011)			
Total SSDI benefits paid (confirmatory)	$7,531	$7,507	$24* ($11)
Number of months with SSDI payments	7.49	7.49	0.00 (<0.01)
Total SSI benefits paid	$339	$338	$0 ($6)
Number of months with SSI payments	1.37	1.38	-0.01 (<0.01)

Source: SSA administrative records for calendar year 2011 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail). Standard errors appear in parentheses. The means and impact estimates are regression-adjusted. Revised impact estimates include a change in two model covariates and use revised administrative analysis weights, as described in Sections A.1 and A.2. See Stapleton et al. (2013) Exhibit 3-1 for initial estimates. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: T1 = 77,115, C1 =891,598.Unweighted sample sizes: T1 = 77,115, C2 =891,598.

*/**/*** 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 (and with no multiple comparisons adjustment).

Exhibit A-4. 2011 Impact Estimates for Subgroups Defined by Duration of SSDI Receipt Based on Revised Specifications and Weights

Outcome	Short Duration			Long Duration			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2011)							
Total earnings	$1,299	$1,331	-$32 ($39)	$1,147	$1,145	$1 ($28)	-$34 ($48)
Employment during year (%)	16.78	16.64	0.14 (0.22)	15.85	15.75	0.10 (0.14)	0.04 (0.26)
Earnings above BYA (%)	2.70	2.73	-0.04 (0.10)	2.31	2.25	0.06 (0.13)	-0.10 (0.17)
Earnings above 2x BYA (%)	1.11	1.21	-0.10 (0.07)	0.88	0.89	-0.01 (0.05)	-0.10 (0.09)
Earnings above 3x BYA (%)	0.68	0.70	-0.02 (0.07)	0.47	0.47	0.00 (0.03)	-0.02 (0.08)
Benefit Outcomes (May–December 2011)							
Total SSDI benefits paid	$8,304	$8,270	$33 ($19)	$7,197	$7,180	$17 ($14)	$17 ($23)
Number of months with SSDI payments	7.57	7.57	0.00 (0.01)	7.46	7.45	0.00 (0.01)	0.00 (0.01)
Total SSI benefits paid	$367	$368	-$1 ($9)	$326	$325	$1 ($7)	-$2 ($12)
Number of months with SSI payments	1.09	1.10	-0.01 (<0.01)	1.49	1.50	-0.01 (<0.01)	0.00 (<0.01)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail). Standard errors appear in parentheses. The means and impact estimates are regression-adjusted. Revised impact estimates include a change in two model covariates and use revised administrative analysis weights, as described in Sections A.1 and A.2. See Stapleton et al. (2013) Exhibit 3-2 for initial estimates. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: Short Duration T1 = 38,669, Short Duration C1 = 209,790, Long Duration T1 = 38,446, Long Duration C2 = 681,808.

*/**/*** 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 (and with no multiple comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit A-5. 2011 Impact Estimates for Subgroups Defined by Baseline SSI Status Based on Revised Specifications and Weights

Outcome	SSDI-Only			Concurrent			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2011)							
Total earnings	$1,299	$1,305	-$6 ($29)	$712	$734	-$21 ($25)	$15 ($38)
Employment during year (%)	16.28	16.12	0.16 (0.13)	15.46	15.57	-0.11 (0.28)	0.27 (0.31)
Earnings above BYA (%)	2.71	2.64	0.06 (0.13)	1.17	1.30	-0.12 (0.10)	0.19 (0.17)
Earnings above 2x BYA (%)	1.12	1.15	-0.03 (0.06)	0.17	0.23	-0.06 (0.05)	0.03 (0.08)
Earnings above 3x BYA (%)	0.63	0.64	-0.01 (0.03)	0.07	0.07	0.00 (0.03)	-0.01 (0.04)
Benefit Outcomes (May–December 2011)							
Total SSDI benefits paid	$8,373	$8,352	$21 ($13)	$3,735	$3,708	$28 ($23)	-$7 ($26)
Number of months with SSDI payments	7.53	7.54	0.00 (<0.01)	7.29	7.27	0.03 (0.03)	-0.03 (0.03)
Total SSI benefits paid	$33	$34	-$1 ($2)	$1,716	$1,709	$7 ($28)	-$9 ($28)
Number of months with SSI payments	0.07	0.07	0.00 (<0.01)	7.25	7.27	-0.02 (0.02)	0.02 (0.03)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail). Standard errors appear in parentheses. The means and impact estimates are regression-adjusted. Revised impact estimates include a change in two model covariates and use revised administrative analysis weights, as described in Sections A.1 and A.2. See Stapleton et al. (2013) Exhibit 3-3 for initial estimates. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: SSDI-only T1 = 64,709, SSDI-only C1 = 694,270, Concurrent T1 = 12,406, Concurrent C1 = 197,328.

*/**/*** 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 (and with no multiple comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit A-6. 2012 Impact Estimates on Earnings and Benefits Outcomes Based on Revised Specifications and Weights

Outcome	T1 Mean	C1 Mean	Impact Estimate
Earnings (January–December 2012)			
Total earnings (confirmatory)	$1,282	$1,274	$8 ($29)
Employment during year (%)	15.39	15.35	0.04 (0.13)
Earnings above BYA (%)	2.68	2.58	0.09 (0.11)
Earnings above 2x BYA (%)	1.11	1.13	-0.02 (0.07)
Earnings above 3x BYA (%)	0.58	0.62	-0.04 (0.03)
Benefit Receipt (January–December 2012)			
Total SSDI benefits paid (confirmatory)	$11,391	$11,322	$70*** ($17)
Number of months with SSDI payments	11.03	10.97	0.05*** (0.01)
Total SSI benefits paid	$457	$460	-$3 ($8)
Number of months with SSI payments	1.98	1.99	-0.01 (0.01)

Source: SSA administrative records for calendar year 2012 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail). Standard errors appear in parentheses. The means and impact estimates are regression-adjusted. Revised impact estimates include a change in two model covariates and use revised administrative analysis weights, as described in Sections A.1 and A.2. See Stapleton et al. (2014) Exhibit 3-1 for initial estimates. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: T1 = 77,115, C1 =891,598.

*/**/*** 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 (and with no multiple comparisons adjustment).

Exhibit A-7. 2012 Impact Estimates for Subgroups Defined by Duration of SSDI Receipt Based on Revised Specifications and Weights

Outcome	Short-Duration			Long-Duration			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2012)							
Total earnings	$1,417	$1,412	$5 ($43)	$1,223	$1,214	$9 ($32)	-$4 ($54)
Employment during year (%)	15.59	15.49	0.09 (0.20)	15.31	15.28	0.02 (0.17)	0.07 (0.26)
Earnings above BYA (%)	3.03	2.94	0.09 (0.13)	2.53	2.43	0.10 (0.12)	0.00 (0.18)
Earnings above 2x BYA (%)	1.20	1.35	-0.15* (0.07)	1.07	1.03	0.04 (0.08)	-0.18* (0.11)
Earnings above 3x BYA (%)	0.74	0.78	-0.04 (0.07)	0.51	0.56	-0.05 (0.03)	0.01 (0.08)
Benefit Outcomes (January–December 2012)							
Total SSDI benefits paid	$12,314	$12,271	$42* ($23)	$10,993	$10,918	$75*** ($23)	-$32 ($32)
Number of months with SSDI payments	11.08	11.05	0.04* (0.02)	11.00	10.94	0.06*** (0.02)	-0.03 (0.03)
Total SSI benefits paid	$403	$408	-$5 ($10)	$480	$482	-$1 ($10)	-$4 ($14)
Number of months with SSI payments	1.55	1.57	-0.02 (0.01)	2.17	2.17	0.00 (0.01)	-0.02 (0.02)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail). Standard errors appear in parentheses. The means and impact estimates are regression-adjusted. Revised impact estimates include a change in two model covariates and use revised administrative analysis weights, as described in Sections A.1 and A.2. See Stapleton et al. (2014) Exhibit 3-2 for initial estimates. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: Short Duration T1 = 38,669, Short Duration C1 = 209,790, Long Duration T1 = 38,446, Long Duration C2 = 681,808.

*/**/*** 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 (and with no multiple comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit A-8. 2012 Impact Estimates for Subgroups Defined by Baseline SSI Status Based on Revised Specifications and Weights

Outcome	SSDI-Only			Concurrent			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2012)							
Total earnings	$1,390	$1,379	$10 ($36)	$796	$800	-$4 ($32)	$15 ($48)
Employment during year (%)	15.54	15.38	0.17 (0.14)	14.70	15.21	-0.51 (0.36)	0.67* (0.39)
Earnings above BYA (%)	2.92	2.80	0.12 (0.12)	1.59	1.59	0.00 (0.12)	0.11 (0.17)
Earnings above 2x BYA (%)	1.30	1.30	-0.01 (0.09)	0.26	0.34	-0.08 (0.06)	0.08 (0.11)
Earnings above 3x BYA (%)	0.68	0.74	-0.06 (0.04)	0.13	0.11	0.02 (0.06)	-0.07 (0.07)
Benefit Outcomes (January–December 2012)							
Total SSDI benefits paid	$12,645	$12,582	$63*** ($19)	$5,741	$5,667	$74 ($49)	-$11 ($53)
Number of months with SSDI payments	11.09	11.04	0.05** (0.01)	10.76	10.68	0.09* (0.04)	-0.04 (0.05)
Total SSI benefits paid	$36	$37	-$1 ($3)	$2,356	$2,367	-$11 ($39)	$11 ($39)
Number of months with SSI payments	0.13	0.14	0.00 (<0.01)	10.33	10.34	-0.01 (0.04)	0.01 (0.04)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail). Standard errors appear in parentheses. The means and impact estimates are regression-adjusted. Revised impact estimates include a change in two model covariates and use revised administrative analysis weights, as described in Sections A.1 and A.2. See Stapleton et al. (2014) Exhibit 3-3 for initial estimates. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: SSDI-only T1 = 64,709, SSDI-only C1 = 694,270, Concurrent T1 = 12,406, Concurrent C1 = 197,328.

*/**/*** 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 (and with no multiple comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit A-9. 2013 Impact Estimates on Earnings and Benefits Outcomes Based on Revised Specifications and Weights

Outcome	T1 Mean	C1 Mean	Impact Estimate
Earnings (January–December 2013)			
Total earnings (confirmatory)	$1,270	$1,266	$4 ($23)
Employment during year (%)	13.42	13.25	0.17 (0.12)
Earnings above BYA (%)	2.64	2.54	0.10 (0.06)
Earnings above 2x BYA (%)	1.10	1.17	-0.07 (0.04)
Earnings above 3x BYA (%)	0.61	0.65	-0.04 (0.03)
Benefit Receipt (January–December 2013)			
Total SSDI benefits paid (confirmatory)	$11,368	$11,230	$138*** ($22)
Number of months with SSDI payments	10.78	10.69	0.09*** (0.02)
Total SSI benefits paid	$438	$443	-$5 ($8)
Number of months with SSI payments	1.90	1.91	-0.01 (0.01)

Source: SSA administrative records for calendar year 2011 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail). Standard errors appear in parentheses. The means and impact estimates are regression-adjusted. Revised impact estimates include a change in two model covariates and use revised administrative analysis weights, as described in Sections A.1 and A.2. See Wittenburg et al. (2015) Exhibit 3-1 for initial estimates. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: T1 = 77,115, C1 =891,598.

*/**/*** 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 (and with no multiple comparisons adjustment).

Exhibit A-10. 2013 Impact Estimates for Subgroups Defined by Duration of SSDI Receipt Based on Revised Specifications and Weights

Outcome	Short-Duration			Long-Duration			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2013)							
Total earnings	$1,462	$1,476	-$13 ($38)	$1,186	$1,175	$11 ($31)	-$25 ($49)
Employment during year (%)	14.39	14.42	-0.03 (0.20)	13.00	12.75	0.26 (0.16)	-0.29 (0.25)
Earnings above BYA (%)	3.11	3.04	0.07 (0.10)	2.43	2.32	0.11 (0.11)	-0.04 (0.15)
Earnings above 2x BYA (%)	1.31	1.44	-0.13* (0.06)	1.01	1.05	-0.04 (0.05)	-0.09 (0.08)
Earnings above 3x BYA (%)	0.77	0.83	-0.06 (0.06)	0.54	0.57	-0.03 (0.04)	-0.03 (0.07)
Benefit Outcomes (January–December 2013)							
Total SSDI benefits paid	$12,204	$12,105	$99** ($42)	$11,007	$10,859	$147*** ($28)	-$49 ($51)
Number of months with SSDI payments	10.79	10.72	0.07** (0.03)	10.77	10.67	0.10*** (0.02)	-0.03 (0.03)
Total SSI benefits paid	$376	$375	$1 ($8)	$465	$472	-$7 ($10)	$8 ($13)
Number of months with SSI payments	1.49	1.51	-0.02 (0.02)	2.08	2.09	-0.01 (0.01)	-0.02 (0.03)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail). Standard errors appear in parentheses. The means and impact estimates are regression-adjusted. Revised impact estimates include a change in two model covariates and use revised administrative analysis weights, as described in Sections A.1 and A.2. See Wittenburg et al. (2015) Exhibit 3-2 for initial estimates. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: Short Duration T1 = 38,669, Short Duration C1 = 209,790, Long Duration T1 = 38,446, Long Duration C2 = 681,808.

*/**/*** 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 (and with no multiple comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit A-11. 2014 Impact Estimates for Subgroups Defined by Baseline SSI Status Based on Revised Specifications and Weights

Outcome	SSDI-Only			Concurrent			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2013)							
Total earnings	$1,385	$1,383	$2 ($29)	$751	$737	$14 ($30)	-$12 ($42)
Employment during year (%)	13.78	13.60	0.18 (0.14)	11.83	11.70	0.13 (0.28)	0.05 (0.31)
Earnings above BYA (%)	2.90	2.77	0.13 (0.09)	1.44	1.47	-0.03 (0.11)	0.16 (0.14)
Earnings above 2x BYA (%)	1.28	1.35	-0.07 (0.05)	0.30	0.35	-0.05 (0.06)	-0.02 (0.08)
Earnings above 3x BYA (%)	0.71	0.76	-0.05 (0.04)	0.13	0.12	0.01 (0.06)	-0.06 (0.07)
Benefit Outcomes (January–December 2013)							
Total SSDI benefits paid	$12,601	$12,467	$134*** ($25)	$5,810	$5,682	$128** ($53)	$6 ($58)
Number of months with SSDI payments	10.84	10.75	0.09*** (0.02)	10.50	10.42	0.08 (0.05)	0.01 (0.05)
Total SSI benefits paid	$36	$37	-$1 ($2)	$2,250	$2,271	-$21 ($39)	$20 ($40)
Number of months with SSI payments	0.16	0.16	0.00 (<0.01)	9.75	9.81	-0.07 (0.06)	0.07 (0.06)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 for further detail). Standard errors appear in parentheses. The means and impact estimates are regression-adjusted. Revised impact estimates include a change in two model covariates and use revised administrative analysis weights, as described in Sections A.1 and A.2. See Wittenburg et al. (2015) Exhibit 3-3 for initial estimates. Dollar amounts are not adjusted for inflation.

Unweighted sample sizes: SSDI-only T1 = 64,709, SSDI-only C1 = 694,270, Concurrent T1 = 12,406, Concurrent C1 = 197,328.

*/**/*** 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 (and with no multiple comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Appendix B. Impact Methodology for Survey Outcomes

This appendix provides details about the impact methodology for outcomes derived from the Stage 1 36-Month Survey.

B.1 Estimation Procedure

The estimation procedure for impacts on *administrative outcomes* is described in Appendix A of the *Third-Year Snapshot of Earnings and Benefits Impacts for Stage 1*. That procedure includes steps that reduce the burden of computation that stems from the very large sample size and our combination of computationally intensive methods that estimate asymptotically unbiased standard errors[118]:

1) Initial regression in the control group,
2) Calculation of residuals for both treatment and control subjects based on the control group regression coefficients,
3) Formation of random groups for data reduction, and
4) Regressions of average group residuals on treatment status.

The impact estimation procedure for *survey outcomes* dispenses with some of the computation-saving complexities of the methodology for administrative data, which are unnecessary for the much smaller survey analysis sample, but in other respects is the same. (The Stage 1 survey analysis sample has 5,735 subjects compared with the administrative analysis sample of 968,713 subjects.) The Stage 1 procedure for survey outcomes is largely similar to the Stage 2 impact estimation procedure, described in the *2016 Stage 2 Interim Process, Participation, and Impact Report*.

Our basic impact estimation model for survey outcomes is:

$$(1)\quad y_{ij} = \beta_0 + \beta_1 T_{1ij} + X_{ij}\Phi + \varepsilon_{ij}$$

where y_{ij} is a survey outcome measure for beneficiary *i* in site *j* (*j* = 1,2, …, 10),
T_{1ij} is an indicator of whether beneficiary *i* in site *j* has been randomized into the T1 group (= 1 if so, = 0 if in the C1 group),
X_{ij} is a vector of baseline characteristics (listed in Exhibit A-2 of *Third-Year Snapshot of Earnings and Benefits Impacts for Stage 1*) for individual *i* in site *j*,
β_0 is the model intercept,
β_1 is the overall impact of the T1 treatment (vs. no treatment for the C1 group),
Φ is a vector of coefficients, and
ε_{ij} is an error term that is beneficiary- and site-specific (discussed below).

We estimate model (1) by weighted least squares regression, using the SURVEYREG procedure in SAS.

[118] Calculation of multiple comparison adjustments for the two confirmatory outcomes also adds to the computational complexity of estimates for those two outcomes. All of the survey-based outcomes are considered exploratory.

The estimated standard errors are computed using the analog of the "variance stabilization" method that we used for estimation of standard errors for estimates of impacts on outcomes measured in administrative data. Variance stabilization errs on the side of a larger standard error when there is a discrepancy between two different methods for estimating the standard error.[119] We first estimate the model while specifying that the ε_{ij} are correlated across individuals within site and independent across sites (i.e., an "unconditional" standard error that treats the sites as clusters, sometimes called the "cluster-robust" standard error). Next, we estimate the model while specifying that the ε_{ij} are independent between and within sites (i.e., a "conditional" standard error that treats the sites as strata, with sites entering into the model as dummy variables, sometimes called the "robust, unclustered" standard error).[120]

These two standard errors are appropriate for different purposes. The unconditional standard errors are designed to support inferences about what would happen with a national implementation of the BOND benefit offset. In contrast, the conditional standard errors are designed to support inferences about what would happen if one treatment variation or another were implemented throughout the 10 sites only. Standard theoretical statistical analysis implies that the *true* unconditional standard errors are at least as large and usually (often considerably) larger than the conditional standard errors. This is because unconditional inference requires us to extrapolate from the 10 sites to the rest of the nation. However, the *estimated* (not true) unconditional standard errors are noisy (unstable) due to the fact that they use observed variation among a small number of sites—only 10. We stabilize the unconditional standard errors by replacing them with corresponding conditional standard errors whenever the unconditional standard error is smaller than the conditional standard error.[121] Both sets of standard errors are estimated using Taylor series linearization in the SURVEYREG procedure in SAS.

For both the unconditional and the conditional model, we compute the estimated standard error for the estimator of β_1 from the estimated variance-covariance matrix, using the ESTIMATE statement. Following the variance stabilization method, we report the standard error for each estimate that is the maximum of the conditional standard error and the unconditional standard error. The p-value for the t-statistic implied by the impact estimate (always from the unconditional model) and the reported standard

[119] The specific method described here is in the spirit of Hanson (1978) and Wolter (1985), where other variance stabilization methods were used.

[120] When estimating the unconditional standard error, the covariates omit site dummies. When estimating the conditional standard error, the covariates include site dummies.

[121] Our simulations have shown that the likelihood of the conditional standard error being larger than the unconditional standard error increases as the true cross-site variance of impacts decreases. In a simulation of very small true cross-site variance of impacts, we found that the 90% confidence interval contained the true effect 92.3% of the time. This result shows that when true cross-site variance is relatively small (and so occasionally the conditional standard error is larger than the unconditional standard error), the variance stabilization method is conservative, sacrificing some statistical power to avoid displaying grossly inconsistent variance estimates for pairs of statistics where generally similar variances are expected. Given the statistical issues, such conservative inference seems appropriate.

error is calculated using 9 degrees of freedom, regardless of whether the reported standard error is the conditional or unconditional standard error.[122]

All impact estimates for survey outcomes are considered exploratory, and therefore we do not adjust the p-values for these estimates for multiple comparisons.

B.2 Stage 1 36-Month Survey Sample Design

The goal of the Stage 1 36-Month Survey was to interview 80 percent of a sample of 10,000 BOND subjects—obtaining 4,000 completed interviews with T1 subjects to statistically represent the entire T1 group and 4,000 completed interviews with C1 subjects to statistically represent the entire C1 group. Both the C1 and T1 samples were selected to represent the national BOND-eligible population, separately. That is, the C1 sample represents the BOND-eligible population under current law as currently administered, and the T1 sample represents the same population under the offset as implemented under BOND. In addition, we stratified the sample to increase the study's ability to detect impacts on T1 subjects employed at baseline, without unduly reducing its ability to detect impacts for all T1 subjects.

To select the sample, we used a stratified multistage sample design that included the initial selection of BOND sites, selection of the full T1 and C1 groups within the selected sites, and then selection of T1 and C1 subjects for the Stage 1 36-Month Survey. In the final step, we used 11 categorical measures of beneficiary characteristics to stratify the sample, based on cross-classification of all categories.[123] This yielded 1,244,160 possible strata or cells, of which 38,388 were populated. Of the populated cells, some were populated for one arm (T1 or C1), but not both. To address that issue, we sorted and collapsed adjacent cells. To help ensure that adjacent records were similar with respect to as many of the variables as feasible, we implemented serpentine sorting (Chromy 1979; Williams and Chromy 1980) in which the sort order was reversed as each boundary was crossed for higher-level sort variables. Given this sorting, cases within collapsed cells differ by fewer variables than if cases were sorted and collapsed under regular sorting. This procedure yielded 654 cells populated by both T1 and C1 cases.

After collapsing cells, we independently selected a probability sample of T1 cases and a probability sample of C1 cases within each collapsed cell. To account for potential survey nonresponse, we drew a sample larger than our target of 8,000 completed interviews. For the purpose of sample release, we then

[122] It is the national representativeness of the impact results that leads to the use of 9 degrees of freedom in the t-tests. Results that only generalize to the 10 BOND sites would use a number of degrees of freedom based on the number of study subjects in the impact comparison, rather than the number of study sites.

[123] The 11 categorical measures come from administrative data. Ten are baseline characteristics, from the month prior to random assignment; see Exhibit B-4 for details. The eleventh variable, 2012 annual earnings (three levels), is based on 2012 earnings as reported to SSA by the IRS. We selected approximately half of the survey sample from those with 2012 earnings. Although 2012 earnings are potentially endogenous (affected by the treatment and correlated with survey outcome variables), all estimates use weights that eliminate the effect of endogeneity on reported statistics. That follows from the fact that we developed the final survey weights independently for the two samples so that a) weighted estimates of outcomes from the T1 sample are unbiased estimates of outcomes for the national BOND population under the BOND offset treatment, and b) weighted estimates of outcomes from the C1 sample are unbiased estimates of outcomes for the national BOND population under current law.

partitioned this sample into 11 replicate samples. Each replicate was an independent random sample in which the comparability of T1 and C1 samples was preserved. During data collection, we released cases in waves, one replicate per wave. In all, we released 5,133 T1 cases and 5,227 C1 cases (10,360 total). Mathematica completed interviews with 2,916 T1 cases and 2,819 C1 cases.[124] The weighted[125] response rates for the treatment group and control group samples were 59 percent and 58 percent, respectively; the difference is not statistically significant.[126]

The analyses of the Stage 1 36-Month Survey data use the analysis weights developed for the survey and reflect the complex (stratified, multistage) sample design. The weights also include adjustments for nonresponse and raking to match statistics on administrative variables for the weighted survey sample to the corresponding statistics for the full Stage 1 samples for purposes of matching.[127]

B.3 Analysis Weights

B.3.1. Summary

The analysis weights for the Stage 1 36-Month Survey account for the different ways in which the survey respondents were sampled and for differences in the response rates across participant characteristics. The weights were developed separately for the T1 and C1 samples, following our sampling approach. As a final step, we adjusted the weights for each group so that certain weighted baseline statistics based on administrative data for the survey respondents in both groups are the same as comparable benchmarks from administrative data for the population, respectively (raking). Furthermore, we adjusted the weights for each group, separately, so that weighted 2014 statistics based on administrative data for certain earnings and DI benefit outcomes for the survey respondents are the same as comparable statistics from administrative data for all members of the corresponding group. Given the complex randomization procedures and low response rates, the use of Stage 1 36-Month Survey weights is required to minimize possible non-response bias. The use of the weights in conjunction with the method used to estimate standard errors ensures that every reported p-value approximates the true likelihood of finding a value of the test statistics as large, or larger than the observed test statistic if the benefit offset had no impact on the outcome. Similarly, under hypothetical repeated subsampling and re-randomization of the sample, confidence intervals of a specified percentage (for example, 95 percent) would contain the true effect of the benefit offset with the same frequency).

We prepared the Stage 1 36-Month Survey weights as the product of several components, each computed separately, and in a sequential manner that builds on components developed at earlier steps to achieve the desired result at the next step in the process. Here, we describe each of the weight components in terms of

[124] Over 20 percent of the released sample was not located (1,475 beneficiaries), ineligible or unable to participate (611 beneficiaries either had died, or were incarcerated, in an institution or had moved out of the United States).

[125] We used the sampling weight (the inverse of the selection probability) to calculate the weighted response rate. Thus, the weighted response rate reflects the success of data collection with regard to the population sampled.

[126] The p-value for the two-tailed test for equal non-response percentages is 0.45.

[127] After nonresponse adjustment, the weighted distribution of the respondents may no longer align with the distribution of the original population. Raking ensures that marginal distributions of characteristics using weighted data on respondents match the corresponding distributions in the original population.

the objective the component was designed to accomplish, how the component was computed, and the impact each component played in the derivation of the overall weight (as reflected by the amount of variation in the overall weights accounted for by each component or a product of specific components). Using these results we demonstrate that the weighting adjustments are necessary and that their use serves to substantially reduce any bias in the estimates that may have been introduced through the sample process and survey nonresponse, while somewhat reducing the statistical precision (widening the margin of error) in the survey estimates. To minimize any bias in making inferences about effects of the BOND benefit offset and other population or experimental parameters of interest, we use the Stage 1 36-Month Survey analysis weights, and we strongly recommend that others using these data for similar purposes do the same.

The Stage 1 36-Month Survey weights consist of the product of the six components summarized in Exhibit B-1 (for example, the sixth component trims the product of the first five components that formed the initial Stage 1 36-Month Survey weight).

Exhibit B-1. Stage 1 36-Month Survey Weight Components

Component	Purpose	Cases Affected
1. Stage 1 Subject Selection	To account for the creation of the Stage 1 analysis sample (from which the Stage 1 36-Month Survey sample was subsampled)	
Site Selection	To account for probability selection of sampled sites	10 sampled sites out of 53 sites
T1 and C1 Assignment and Sampling	To account for random assignment of cases within 10 selected sites and oversampling of short-duration cases	77,115 T1 and 891,598 C1 cases
2. Survey Selection	To account for T1 and C1 sample selection	12,000 sampled cases
Stratification and oversampling earners	To account for probability selection of T1 and comparable C1 by strata	6,000 T1 and C1 cases each
3. Released Sample	To account for the release to the field of less than 100 percent of the sample replicates	10,360 released cases
4. Survey Nonresponse	Adjusting the respondent weights to reduce potential nonresponse bias	
Non-Located	To account for cases that were released but could not be located	8,885 located cases
Located and eligible but nonresponding cases	To account for cases that were released, located and eligible for the survey but data not completed	5,735 completed (the respondents)
5. Raking	To align the weighted counts to population statistics from administrative data	5,735 respondents to the survey
6. Trimming	To reduce excessive variability in the weights due to weight outliers	5,735 respondents to the survey
7. Re-raking	To re-align the weighted counts to population statistics from administrative data	5,735 respondents to the survey

As summarized in Exhibit B-1, the Stage 1 survey sample weight takes into account multiple components of the process that led to the generation of the survey respondent sample. Component 1 includes the preparation of the list of Stage 1 subjects and accounts for the selection of 10 BOND sites, assignment of T1 and C1 cases and the oversampling of short-duration cases in T1. Component 2 accounts for the selection of T1 and C1 samples from the Stage 1 subject list for the Stage 1 36-Month Survey (6,000 T1 cases and 6,000 C1 cases). Next, because not all sampled cases were released for the survey, we adjusted the weights from component 2 for unreleased sample (component 3). For component 4, we used a two-

step propensity modeling method to adjust the weight for non-located cases (653 out of 5,133 released sample in T1; and 822 out of 5,227 released sample in C1), and for located and eligible but nonresponding cases (1,246 out of 4,162 located and eligible sample in T1; and 1,296 out of 4,115 located and eligible sample in C1). The response propensity models used the baseline covariates (Exhibit B-4)[128] as predictors for the propensity to respond to the survey, separately for treatment and control cases. We then used the estimated propensity scores as the weighting adjustment factors (Little 1992; Iannacchione 2003).

For component 5, we post-stratified or raked the weights so that certain weighted estimates from the survey would match benchmarks from the BOND population. We used a raking technique—iterative proportional fitting (Deming and Stephan 1940)—to adjust the weights so that weighted counts/totals from the sample (T1 and C1 cases separately) align with the national BOND population totals from SSA administrative data.[129] We developed raking cells based on 11 administrative variables: eight baseline (2011) variables that were among the 11 administrative variables used to stratify the sample for selection purposes (those for which values do not change with time) and 2014 values for the three variables that can change with time, based on earnings, DI benefits, and concurrent status. As with the nonresponse adjustments, we performed the raking for treatment and control cases separately. In component 6 we performed weight trimming (Potter 1990) to reduce extreme weights that might create unnecessary variance inflation. Finally, we re-raked the weights to re-align the sample statistics with the same administrative statistics used in component 5.

As Exhibit B-1 indicates, because the Stage 1 36-Month Survey respondents are based on the BOND Stage 1 analysis sample (all of the beneficiaries in the T1 and C1 samples), rather than a direct sample from the national BOND population, the weights must first account for the process that generated the BOND analysis sample. This weight component was created previously following the methods presented in Section B.3 of the *Evaluation Analysis Plan*. As a result this appendix focuses on the remaining components of the process.

---

[128] The sampling and weighting processes used 12 covariates from administrative data. Because one variable—earnings (or earnings group)—can only be accessed by qualified SSA employees, such an SSA employee ran the final programs to construct the weights.

[129] The T1 and C1 respondent samples (including responding ineligibles) were each raked to the same set of population target values for the 8 baseline characteristics. We estimated the baseline 2011 population target values by computing the weighted counts (using the administrative weights) from the combined T1 and C1 administrative samples, for each of the categories for each of the baseline characteristics in Exhibit B-11. Because the administrative weights were designed to weight each of the T1 and C1 administrative samples to totals for the national BOND-eligible population, the totals for the two groups combined are twice as large as the actual population values; hence, we divided the administrative weights by two to produce the target counts for raking. As a result, the characteristics of the T1 and C1 survey respondents and known ineligibles given in Exhibit B-11 for the baseline characteristics, once weighted, are identical to the corresponding characteristics for the population they represented at baseline. We also raked the weights for the T1 and C1 respondents and known ineligibles to three 2014 benefit and earnings outcomes from the administrative data, but in this case we used different totals for T1 and C1. That's because we expected the offset to affect each of these outcomes, so the goal was to produce weighted totals for each survey sample that match the totals for the corresponding sample of all T1 or C1 subjects.

Because some of the covariates used to create the weights involve IRS earnings data (which are only accessible to qualified SSA staff), Mathematica designed programs to create the weights and submitted them to SSA for execution. Because the weights rely on the earnings data, they are not included in the survey data file; only a qualified SSA employee may access the weights.

Section B.3.2 reviews background on the implementation of BOND that is relevant to the construction of the weights. Section B.3.3 presents the specific procedures associated with the computation of each of the weight components. Section B.3.4 provides instructions on the use of these weights for analytical purposes.

B.3.2. Background

We designed the Stage 1 36-Month Survey to support estimates of impacts on the national BOND-eligible beneficiary population for outcomes not reported in administrative data for the Stage 1 analysis sample. The Stage 1 analysis sample is itself a probability of the national BOND-eligible population at the time of random assignment. Hence, the weights for the survey need to reflect both the selection methodology for the Stage 1 analysis sample as well as the methodology for selecting the survey sample and survey non-response.

The selection of the full Stage 1 BOND sample was a multistep process. The first step was site selection. To limit the size of the study, we conducted the BOND demonstration only in 10 randomly selected geographical areas or sites.[130] We selected 10 sites using a stratified sampling method that makes it possible to generalize the study results to all BOND-eligible beneficiaries nationwide. Site selection used probability proportional to site size based on the number of beneficiaries included in each site. We placed the sites into eight strata reflecting a combination of census region membership (Northeast, Midwest, South, or West) and the extent (high or low) to which beneficiaries in each site had access to health insurance coverage under the Medicaid Buy-in (MBI) program. Exhibit B-2 provides the stratification of Primary Sample Units (PSUs); each site is a PSU for site selection.

[130] The sites were defined by the coverage areas of the 54 SSA area offices, from which 10 were selected.

Exhibit B-2. Selected Area Offices in the BOND Sample, by Census Region and Proportion of Beneficiaries Living in Medicaid Buy-in States

Stratum	Census Region	Proportion of Beneficiaries in MBI States	Selected SSA Area Offices
1	Northeast	Low	Northern New England
2		High	Western New York
3	Midwest	Low	Greater Detroit
4		High	Wisconsin
5	South	Low	Alabama
			South Florida
			Greater Houston
6		High	DC Metro
7	West	Low	Colorado,
			Wyoming
8		High	Arizona, Southeast California

We followed a multi-step process to select the Stage 1 BOND sample subjects from beneficiaries residing in these three sites. First, we randomly assigned the sampling frame of all BOND-eligible beneficiaries[131] in the 10 selected sites to three pools, including a Stage 1 treatment pool (T1) that received the offset, a control group that did not (C1), and a Stage 2 solicitation pool. The combined T1 and C1 pools formed the analysis sample frame for the Stage 1 evaluation. Because the Stage 2 solicitation pool excluded concurrent cases, the C1 sample includes a disproportionately large number of concurrent cases. Furthermore, assignments to T1 oversampled short-term cases so that the pool would contain an approximately equal number of short-term and long-term beneficiaries, to support short-term/long-term comparisons.[132] This assignment scheme implies that the C1 sample also has a disproportionately large number of long-term cases. The first component of the Stage 1 36-Month Survey weights adjust for these differences. With those weights, statistics based on weighted *administrative data* from both the T1 and C1 samples are unbiased estimates for the national BOND-eligible population.

After we had selected the Stage 1 sample, we discovered that a small number of beneficiaries were in multi-subject households, and that some of these beneficiaries were assigned to a different BOND group than one or more other subjects in the same household—for instance, one might have been assigned to T1 and the second to C1 or the Stage 2 recruitment pool. Multi-beneficiary households arise primarily because some disabled adult child (DAC) beneficiaries receive benefits under the record of a parent who is a disabled worker. Less common cases arose because of DAC siblings or because of a disabled widow beneficiary and a DAC eligible as survivors of a deceased spouse/parent.

Given concerns of cross-over effects that might occur when members of the same household are assigned to different groups, we dropped cases of subjects in households containing another subject assigned to a

[131] To be BOND eligible, an SSDI beneficiary had to be between the ages of 20 and 59 at the time of enrollment and not participating in other Social Security Administration (SSA) demonstration projects.

[132] In the population, approximately one-quarter of the beneficiaries were short-duration beneficiaries.

different group. For estimation of impacts on outcomes measured in administrative data, we reweighted the T1 and C1 sample members in the multiple-subject households to represent all population subjects in multiple-subject households. For purposes of the survey sampling and survey weights, however, we ignored the presence of another subject in the same household. In essence, we are assuming that the behavior of a BOND subject is not affected by the behavior of another BOND subject in the same household if the other subject was assigned to the same group, holding the values of the control variables constant. The number of such subjects in the T1 survey sample is very small (Exhibit B-3). That's because the probability of any beneficiary being assigned to T1 was small, so the probably that two subjects in the same household would be assigned to T1 is tiny. The T1 respondent sample includes only three such subjects. The number of such respondents in the C1 sample is substantially larger, 84, reflecting the higher likelihood that two subjects in the same household would be assigned to C1. If, alternatively, we had given the respondents in each survey sample sufficient weight to represent the entire multi-household beneficiary population, the weights for the three T1 respondents in this groups would have been so high that they would have greatly reduced precision. We instead include a "multi-subject household" indicator (MPHH) as a control variable in the impact analysis, along with many other controls, to minimize possible bias that might arise because the number of such subjects in the C1 sample is higher than in the T1 sample. The possibility of some bias remains, because mean impacts for subjects in such households might be different for those in other households, after controlling for other characteristics. However, the presence of just three such subjects in the T1 sample prevents us from conducting a meaningful test of the hypothesis that the impacts for such subjects are no different than for others. See further discussion in Section B.3.4.

Exhibit B-3. Survey Subjects in Multi-Subject Households

Sample Group	Multi-subject Household Status	Sample Count	Respondent Count
T1	Yes	6	3
	No	5,994	2,913
	Total	6,000	2,916
C1	Yes	171	84
	No	5,829	2,735
	Total	6,000	2,819

Selection of the survey sample from the Stage 1 BOND sample was also a multi-step process. To be eligible for the Stage 1 survey, T1 and C1 subjects must have been alive, in the United States, and noninstitutionalized at the time of data collection. We independently selected the subsets of T1 and C1 cases to be surveyed by a process that ensures that each sample can be used to efficiently estimate statistics for the national BOND population—national implementation of the BOND benefit offset for T1 and under current law for C1. To reduce chance differences between the two samples and between each sample and the national BOND population, we used a common set of strata in the sampling processes. As a result the selected T1 and C1 cases in each stratum shared a set of common characteristics. In addition, to increase the proportion of respondents who were employed at the time of their interview, we oversampled strata that included beneficiaries with positive earnings in 2012. Past research on employment of DI beneficiaries has found that, of all information in SSA administrative records, earnings in any year is the best predictor of earnings in subsequent years.

This stratification plan created more than 650 strata/sampling cells based on the 11 characteristics reported in Exhibit B-4.[133] As a result, each stratum contains T1 and C1 cases that shared a set of common characteristics. We then independently selected probability samples of T1 and C1 within each stratum. For each group, we allocated the sample in each stratum proportional to the size of the stratum in the sampling frame. We then oversampled cases with positive 2012 earnings using the following probabilities of selection, which are based on the use of probability proportionate-to-size (PPS) methods for each case within a given stratum and sample type:

$$P(S)_{h,i} = n(S)_h \frac{MOS(S)_{h,i}}{\sum_{j=1}^{N(S)_h} MOS(S)_{h,j}} \quad ,$$

where:

- $P(S)_{h,i}$ denotes the probability of selecting a given beneficiary eligible for stage 1 sampling (case *i*), from stratum *h*, *h* = 1,…,654, for sample type *S* = T1 or C1.
- $N(S)_h$ denotes the total number of cases in sample type S and stratum h.
- $n(S)_h$ denotes the total number of cases in $N(S)_h$ allocated/selected in sample type *S*, and stratum *h*.
- $MOS(S)_{h,i}$ denotes the measure of size (MOS), for case *i* in stratum *h*. This was set to a value of 1 for cases without 2012 earnings, and a value of 6 for cases with 2012 earnings to oversample cases with earnings in each stratum and sample.

With proportional allocation of the sample to each stratum, we allocated the sample size in each strata, as follows, to meet the overall T1 and C1 sample size requirements n(S):

$$n(S)_h = n(S) \times \frac{\sum_{i=1}^{N(S)_h} MOS(S)_{h,i}}{\sum_{h=1}^{654} \sum_{i=1}^{N(S)_h} MOS(S)_{h,i}} .$$

Once the samples of T1 and C1 cases were selected, we randomly subdivided each of these samples into 11 waves for sequential release.

After data collection, we adjusted the Stage 1 36th-month survey weights to account for unreleased sample waves as well as for survey nonresponse, post-stratification and trimming, as will be discussed in section II.

[133] We used the 11 categorical measures of beneficiary characteristics to stratify the sample based on cross-classification of all categories. This yielded 1,244,160 possible strata or cells, of which 38,388 were populated. Of the populated cells, some were populated for one arm (T1 or C1), but not both. To address that issue, we sorted and collapsed adjacent cells. To help ensure that adjacent records were similar with respect to as many of the variables as possible, we implemented serpentine sorting (Chromy 1979; Williams and Chromy 1980) in which the sort order was reversed as each boundary was crossed for higher-level sort variables. Given this sorting, cases within collapsed cells differ by fewer variables than if we had sorted and collapsed cases under regular sorting. This procedure yielded 654 cells populated by both T1 and C1 cases.

Exhibit B-4. Variables Used to Stratify the Stage 1 Sample for Survey Sample Selection

Variable Name	Year of Data	Category
Site Name (10 level)	All BOND years	Alabama, Arizona/Southeast California, Colorado/Wyoming, DC Metro, Greater Detroit, Greater Houston, Northern New England, South Florida, Western New York, Wisconsin
Age Category (3 level)	2011 (baseline)	< 40, 40 - < 50, 50 +
Gender	2011 (baseline)	Female, Male
Benefit duration status	2011 (baseline)	Short duration, Long duration
Beneficiary type	2011 (baseline)	Disabled worker, disabled adult child, disabled widow(er)
Representative payee	2011 (baseline)	Yes, No
Average indexed monthly earnings tercile	2011 (baseline)	Low, Middle, High
Primary impairment	2011 (baseline)	Back and musculoskeletal, Circulatory system disorders, Digestive system, Genitourinary system disorder, Injuries, Mental disorders, Neoplasms, Nervous system disorders, Other impairments, Respiratory, Severe visual impairments, Unknown impairments
Concurrent	2011 (baseline)	Yes, No
Monthly benefit amount tercile	2012	Lower, Mid, Higher
Earnings group	2012	Zero earning, $1 - $5,000, $5,000 +

B.3.3. Detailed Methodology for the Computation of the Stage 1 36-Month Survey Weights

This section provides details on each of the seven components in the development of the weights (i.e., the rows of Exhibit B-1).

Component 1: Stage 1 Analysis Sample

From Section B.3 of the Evaluation Analysis Plan, the starting distribution of the survey weights based on component 1 were as follows:

$$Component\ 1 = \frac{1}{p_{site}} \times \frac{1}{p_g} \qquad (1),$$

Where p_{site} is the probability that the site is selected and p_g is the probability of that a beneficiary in the site is assigned to the group g (T1 or C1). Exhibit B-5 provides statistics on the distributions of the weights for Component 1. The results show that the weighted count of cases in the administrative group of T1 and C1 cases both sum to the same total of 6,454,576 because the T1 and C1 groups each represent, once weighted, identical populations. Additional information on the creation of the component 1 weights is provided in Appendix A.

Exhibit B-5. Distribution of the Weight Component 1[134]

Sample Group	Cases (n)	Mean	Min	Max	CV (%)	Sum
T1	77,115	83.70	18.61	3391.63	90.36	6,454,576
C1	891,598	7.23	2.17	129.64	44.18	6,454,576

Component 2: Survey Selection

We calculated the sampling weight of the Stage 1 36-month survey sample as the inverse of the selection probability to be in the sample, conditional on being in the Stage 1 Sample for group g. That is,

$$Sampling\ weight = \frac{1}{p_{g,survey}} \qquad (2),$$

The distribution of this sampling weight is given in Exhibit B-6.

Exhibit B-6. Distribution of the Sampling Weight

Sample Group	Cases (n)	Mean	Min	Max	CV (%)	Sum
T1	6,000	12.69	2.31	54.75	72.65	76,150
C1	6,000	149.89	27.81	570.08	70.76	899,311

Then we multiplied this sampling weight with the weight from Component 1 to get the Component 2 weight:

$$Component\ 2 = Component\ 1\ \times \frac{1}{p_{g,survey}} \qquad (3).$$

The sum of these weights across 6,000 Stage 1 36-Month Survey sample members in each group—whether actually released or not—estimates the BOND-eligible SSDI population (around 6.5 million). Statistics for their distributions appear in Exhibit B-7. We note that the sum of the component 2 weights in Exhibit B-7 (6,347,460 for T1 and 6,547,511 for C1) which are the product of component 1 and the sampling weight for the survey sampled cases as shown in (3) are not the same as the sum of the component 1 weights in Exhibit B-5 (at 6,454,576 for both T1 and C1). Likewise, the sum of the sampling weights in Exhibit B-6 is not the same as the number of cases in Exhibit B-5. The reason is that probabilities of selection for each of the cases in the sampling frame of 77,115 T1 cases and 891,598 C1 cases were different because of differences in the survey sample selection methods.[135] As a result the sum of the weights will vary from one random sample to the next, depending on which cases are actually selected. We adjusted the weights in the raking step, component 5, so that the survey weights for the survey respondents and known ineligibles sum to the same total that appears in Exhibit B-6, 6,454,576 (see Exhibit B-12).

[134] These weights were computed previously during selection of 10 BOND sites, assignment of T1 and C1 cases and the oversampling of short-duration cases.

[135] As noted in section B.3.2, we used more than 650 sampling cells, and within each cell we implemented probability proportional to size where the measure of size (MOS) is simply one for cases without earnings, and six for cases with positive 2012 earnings.

Exhibit B-7. Distribution of Component 2

Sample Group	Cases (n)	Mean	Min	Max	CV (%)	Sum
T1	6,000	1,057.91	58.05	60,201.35	130.46	6,347,460
C1	6,000	1,091.25	68.22	12,295.37	92.94	6,547,511

Component 3: Adjusting for Sample Release (Unused Sample)

Because we did not release all sample replicates, Component 3 adjusts the weights in Component 2 for non-released sample waves by a simple ratio adjustment technique within each sampling stratum, separately by treatment and control. In this procedure, we calculated the ratio adjustment factor as a weighted ratio of the sum of weights from all sampled cases to the sum of weights from the randomly released sampled cases.

Given the analytical objectives for the BOND Stage 1 36-Month Survey data, the sample size calculation suggested a balanced sample of 5,000 cases from each of T1 and C1, with a target response rate of 80 percent (*Evaluation Analysis Plan*, p. 53*)*. To anticipate additional nonresponse, however, we selected extra sample of 1,000 cases in each group; that is, we selected 6,000 T1 cases and 6,000 C1 cases initially. During data collection, however, budget and time constraints prevented the release of some of the 12,000 sampled cases. Instead only 10,360 samples were released for the survey (5,133 T1 and 5,227 C1 cases).[136] To bring the weights of the released sample back to the population, it was necessary to adjust their base sampling weights. This adjustment was done within each sampling stratum and separately for treatment and control groups, as follows:

$$Released\ weight = Component\ 2\ \times release\ adjustment\ factor,$$

where the release adjustment factor (RAF_{ki}) for sample member *i* in group *k* (*k* is either treatment or control) is calculated as the ratio of the weighted count from the initial sample to the weighted count from the released sample only

$$RAF_k = \frac{\sum_i^{all\ sample} base\ weight_{ki}}{\sum_i^{released\ sample} base\ weight_{ki}} \qquad (3).$$

Statistics for Component 3 appear in Exhibit B-8.

Exhibit B-8. Distribution of the Weights after Release Adjustment

Sample Group	Cases (n)	Mean	Min	Max	CV (%)	Sum
T1	5,133	1,236.60	58.05	61,302.68	128.86	6,347,460
C1	5,227	1,252.63	75.58	13,451.82	94.55	6,547,511

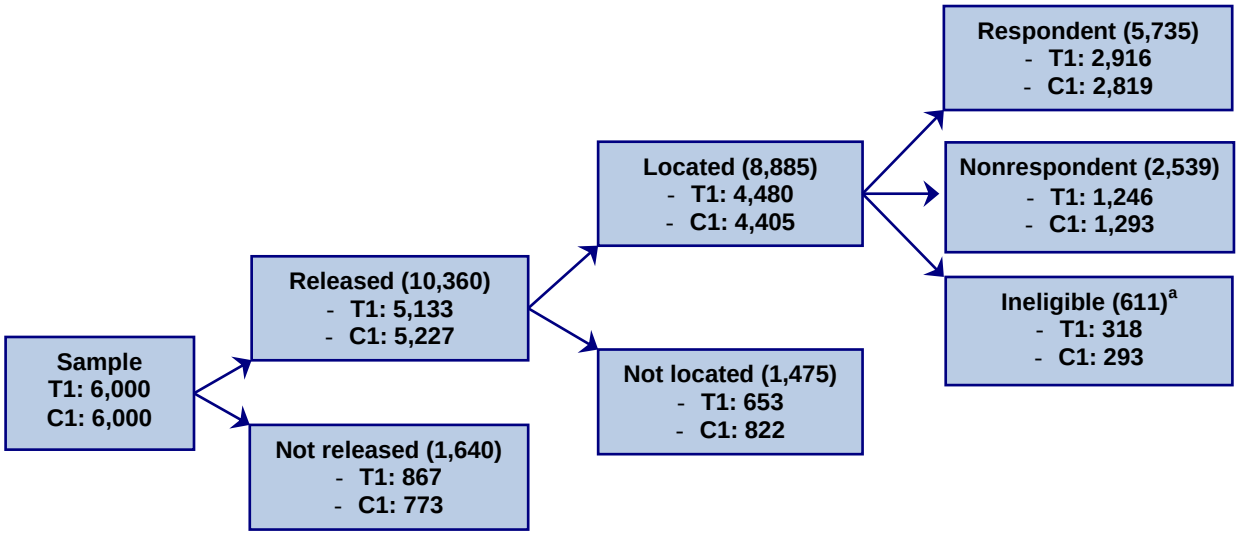
[136] During data collection, Mathematica monitored the overall response rate as well as the response rate by the 12 baseline covariates. Though the final response rate did not meet the targeted 80 percent, we were able to obtain the response rates that are balanced across these 12 baseline covariates.

The adjustment factor ranges from 1.00 to 15.00 in T1, and from 1.00 to 21.20 in C1. We set the adjustment factor (and therefore the release-adjusted weight) for unreleased sample to be zero.

Component 4: Survey Nonresponse

Among the several reasons for sample nonresponse are the inability to locate the sampled beneficiary, lack of contact with cases that could be located, and failure of the contacted beneficiary to complete the survey (partial nonresponse). Partial nonresponse has its own causes; it occurs when the respondent (1) answers some survey items but not all of the critical items or (2) completes part of the survey but breaks off at some point—that is, stops the interview. If any of the critical complete items on the questionnaire are incomplete, we reclassify a partial response as a unit nonresponse and treat it accordingly. Otherwise, we include the partial response as a completed interview. At the end of data collection, the survey team determined the final survey disposition status for each sample member. Accordingly, all sampled beneficiaries were classified by their response status as follows: (1) not released; (2) released, not located; (3) released, located, ineligible; (4) released, located, eligible nonrespondent; (5) released, located, eligible respondent. The number of cases in each group is given in Exhibit B-9.

Exhibit B-9. Sample Disposition by Locating and Eligibility Status



[a]We identified ineligible cases at the time of data collection and excluded 532 beneficiaries who had died, 66 who were incarcerated, 10 who were institutionalized, and 3 who had moved out of the United States.

The final unweighted overall response rate was 58.8 percent, and the weighted overall response rate was 58.0 percent.[137] The response rates were fairly constant across the 12 characteristics used to stratify the sample.

Given that the reasons for nonresponse vary and that the cases within each status may be characterized differently, after the sample release adjustment we performed a two-step weighting adjustment to create Component 4, sequentially to account for different types of unit nonresponse: (1) adjustment for nonlocated sample members, and (2) adjustment for sample members located and eligible, but not responding to the survey,

$$Component\ 4 = Released\ weight\ \times (Nonlocated\ Adjustment) \times (Located, Nonresponse\ Adjustment).$$

Each of these two factors is discussed separately below.

Adjustment for nonlocated sample

We adjusted the release-adjusted weights for the located sample beneficiaries (8,885 cases) to account for nonlocated sample beneficiaries (1,475 cases). We used a logistic regression model using baseline covariates to estimate location propensity for each beneficiary, and adjusted the weight of each located case by dividing it by the estimated response propensity. We performed this weighting-adjustment process separately for T1 and C1.

In determining which variables were significant predictors of located status, we implemented two steps:

First, we ran a chi-squared automatic interaction detector (CHAID) analysis in SPSS to find possibly significant interaction terms. CHAID is normally attributed to Kass (1980) and Biggs et al. (1991), and its application in SPSS is described in Magidson (1993). The CHAID procedure iteratively segments a data set into mutually exclusive subgroups that share similar characteristics based on their effect on nominal or ordinal dependent variables. It automatically checks all variables in the data set and creates a hierarchy that shows all statistically significant subgroups. The algorithm finds splits in the population that are as different as possible based on a chi-square statistic. CHAID is a forward stepwise procedure; it finds the most diverse subgroup, and then further splits each of these subgroups into more diverse sub-subgroups. Sample-size limitations are set to avoid generating cells with small counts, as are limitations on the number of levels of branching. CHAID stops when splits no longer are significant—that is, when the group is homogeneous with respect to variables not yet used or when the cells contain too few cases. The CHAID procedure results in a tree that identifies the set of interactions among the variables' values that have an association with a sample member's propensity to be located.

[137] We calculated these response rates using the American Association for Public Opinion Research (AAPOR) formula for Response Rate 2 (http://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/ResponseRateCalculatorVer3-1_11_22_10.xls), where the denominator includes the cases with unknown eligibility and treats them all as eligible. The response rate using AAPOR Response Rate 4, where the denominator does include the estimate of eligible cases out of the cases with unknown eligibility, is 59.4 percent unweighted and 58.8 percent weighted. The weighted response rates were calculated using the unconditional sampling weights.

Second, given the identified significant interactions from CHAID, we sought to further refine the candidate variables and interaction terms by processing the variables and interactions identified with CHAID through the stepwise selection (using the SAS logistic procedure, unweighted). We used a logistic propensity model, where the response variable was a binary indicator of located status. We evaluated a series of models to determine the final model with the best fit and other properties.[138] After identifying a smaller pool of main effects and interactions for potential inclusion in the final model, we estimated propensities to be located and used them as weighting adjustment factors.

The weights for located beneficiaries were adjusted as follows:

$$Located\ Adjustment = \frac{Released\ weight}{Estimated\ located\ propensity} \quad (4).$$

The estimated propensity of being located ranged from 0.636 to 0.997 (from 0.636 to 0.997 for T1, and from 0.671 to 0.949 for C1). We set the weight for non-located beneficiaries to zero.

Adjustment for located sample not responding

The located sample members include some nonrespondents due to ineligibility, no contact, refusal, and so forth. Hence, we adjusted the weights for eligible respondents (5,735 cases) to account for eligible non-respondents (2,539 cases). We did not include ineligible cases in this process. We used the same approach for located non-respondents as we did for the located adjustment. We used a logistic model to determine the baseline variables that are significant predictors of propensity to be a(n) (eligible) respondent, where the interaction terms were identified through CHAID analysis.

We then estimated the propensity to respond and used it as an adjustment factor for each survey respondent, computing the adjusted weight as follows:

$$Response\ Adjustment = \frac{located\ adjustment}{estimate\ response\ propensity} \quad (5).$$

The estimated propensity of response ranged from 0.497 to 0.874, and the ranges for each group are very similar to each other (from 0.543 to 0.862 for T1, and from 0.497 to 0.874 for C1). We set the weight for nonresponding beneficiaries to zero. We kept the weights of ineligible cases as their located weight; that is, the sum of the weights for the ineligible cases for each group estimates the number of ineligible beneficiaries in the nationwide baseline BOND population. The Component 4 weights for respondents in the located sample appear in the bottom half of Exhibit B-10.

[138] Given the selected variables from the stepwise procedure, we evaluated the goodness of fit of the model using the Hosmer-Lemeshow goodness of fit test, as well as evaluated the association of the predicted probabilities and the observed responses through the magnitude of percent concordant and discordant. The final model for treatment group has a p-value for the Hosmer-Lemeshow goodness of fit test of 0.6607, and percentage concordant and discordant, respectively, equal to 63.4 and 35.0. The final model for the treatment group has a p-value for the Hosmer-Lemeshow goodness of fit test of 0.9082, and percentage concordant and discordant, respectively, equal to 60.6 and 37.6.

Exhibit B-10. Distribution of the Weight Component 4 Factors

Factor	Cases (n)	Mean	Min	Max	CV (%)	Sum
Located sample weights after adjustment for non-location						
T1	4,480	1,418.50	65.37	74,266.85	133.58	6,354,858
C1	4,405	1,497.18	87.71	14,966.28	92.42	6,595,098
Respondent sample weights after adjustment for non-response in located sample*						
T1	2,916	1,889.38	92.39	53,955.47	116.44	5,509,443
C1	2,819	2,109.29	111.57	29,002.17	98.34	5,946,081

*These represent the eligible population among the 6.3 and 6.6 million, respectively, in the T1 and C1 above.

Component 5: Raking

To enhance the accuracy of the estimates, we used a raking procedure to ensure that respondents' adjusted weights conformed to the marginal distributions of the auxiliary variables (Deming and Stephan 1940). The objective of a post-stratification process is to ensure that distributions (or counts) of auxiliary variables, once weighted by the base weights and adjusted for nonresponse, generate estimates that match those known for a given level for the T1 and C1 respondents, separately. In this case, we benchmarked the T1 and C1 respondents' weights back to statistics for the national BOND population from administrative data.

We used a raking procedure that involves an iterative adjustment of the weights, implemented with an iterative proportional fitting algorithm. Under general conditions, the algorithm converges to a solution. Cases included in the raking process were the respondents and ineligible cases only.[139] The domains used for raking include eight baseline domains that were also used for stratification of the Stage 1 BOND sample for purposes of survey sampling (domains that are fixed over the sample period) and three benefit and earnings domains for 2014, to match as closely as feasible to administrative information on these important outcomes (see Exhibit B-11). The C1 and T1 respondent samples were each raked to a common set of values for the eight baseline domains (estimates for the baseline BOND-eligible population based on administrative weights for the full samples), but were raked to different values for the three 2014 domains (estimates based on 2014 values for their respective administrative samples, again using the administrative weights).

[139] The weights for ineligible cases that have been adjusted up to non-locate adjustment and weights for eligible cases that have been adjusted up to non-response adjustment should sum to the population control totals.

Exhibit B-11. Raking Domains: Site, 2011 Baseline Covariates, and 2014 Benefits and Earnings

Variable Name	Year of Data	Category
Site Name (10 level)	All BOND years	Alabama, Arizona/Southeast California, Colorado/Wyoming, DC Metro, Greater Detroit, Greater Houston, Northern New England, South Florida, Western New York, Wisconsin
Age Category (3 level)	2011 (baseline)	< 40, 40 - < 50, 50 +
Sex	2011 (baseline)	Female, Male
Benefit duration status	2011 (baseline)	Short duration, Long duration
Beneficiary type	2011 (baseline)	Disabled worker, disabled adult child, disabled widow(er)
Representative payee	2011 (baseline)	Yes, No
Average indexed monthly earnings tercile	2011 (baseline)	Lower, Mid, Higher
Primary impairment	2011 (baseline)	Back and musculoskeletal, Circulatory system disorders, Digestive system, Genitourinary system disorder, Injuries, Mental disorders, Neoplasms, Nervous system disorders, Other impairments, Respiratory, Severe visual impairments, Unknown impairments
Monthly benefit amount tercile	2014	Lower, Mid, Higher
Concurrent	2014	Yes, No
Earnings group	2014	Zero earning, $1 - $5,000, $5,000 +

The weighted counts based on raked weights should match with the population marginal control totals for these variables. Raking minimized the distance between the weighted sum and reference total count within each cross-classification cell. The resultant adjusted weight (raked weight) was calculated as

$$Component\ 5 = respondent\ weight \times raking\ adjustment\ factor.$$

The raking adjustment factors ranged from 0.428 to 2.515 in T1 and from 0.503 to 1.864 in C1. Statistics for the Component 5 weights appear in Exhibit B-12.

Exhibit B-12. Distribution of Weight Component 5

Component 5	Cases (n)	Mean	Min	Max	CV (%)	Sum
T1	3,234	1,995.85	40.58	49,531/22	118.66	6,454,576
C1	3,112	2,074.09	88.37	26,252.60	95.91	6,454,576

Component 6: Weight Trimming

We evaluated the raked weights for outlier weights, paying particular attention to the weight inflation due to nonresponse adjustments. When the response propensity or response rate is small for some particular characteristics of the sample, this usually leads to large nonresponse adjustment which increase the weights significantly. These nonresponse adjusted weights might reduce potential nonresponse bias, but at the expense of higher sampling error may if the range of weights is large. That is, there is a tradeoff between bias and precision, and when a few weights are extremely large, the mean square error (the sum of the square of the estimator's bias and the estimator's variance) can be reduced by trimming the weights, even if there is a small increase in bias.

Weight trimming is a method used to reduce extreme weights. We trimmed outlier weights within the domains defined by earnings group (no earnings; $1-$5,000; $5,000+) separately for treatment and control groups. Within each trimming domain, we identified extreme weights if they were further than five standard deviations from the mean; that is, above the cut-off point defined by $\overline{w}_d + 5 \times sd_d(w)$, where $\overline{w}_d$ is the mean of raked weights in domain d, and $sd_d(w)$ is the standard deviation of raked weights in domain d. We trimmed extreme weights to the cut-off point. The trimming adjustment factor for record l in domain $d(l)$ is calculated as follows:

$$ADJF\_TRM_l = \begin{cases} \dfrac{\overline{w}_{d(l)} + 5 \times sd_{d(l)}(w)}{raked\ weight_l} & \text{if } raked\ weight_l \geq \overline{w}_{d(l)} + 5 \times sd_{d(l)}(w), \\ 1 & \text{if } raked\ weight_l < \overline{w}_{d(l)} + 5 \times sd_{d(l)}(w). \end{cases}$$

The trimmed-adjusted weight is then the product of the trimmed factor for extreme weights and the raking-adjusted weight, as follows:

$$trimmed\ weight_l = ADJF\_TRM_l \times raked\ weight_l.$$

This procedure resulted in weight adjustment for 26 observations, with adjustment factors ranging from 1.0073 to 1.0174 in T1, and from 1.0043 to 1.0319 in C1). We then redistributed the amount trimmed among other sample members in the same domain in proportion to their untrimmed weights.

Exhibit B-13 provides a summary of the pre-trimmed survey weights developed from the product of components 1-5, by treatment-control and earning groups. Exhibit B-14 provides the summary of the weights after trimming.

Exhibit B-13. Summary of the Weights Before Trimming

Domain	Cases (n)	Mean	Min	Max	CV (%)	Sum	5 × Standard Deviation
T1							
No earnings	2,163	2,590.65	40.58	49,531.22	101.33	5,603,575.73	15,719.69
$1-$5,000	428	893.58	86.21	8,019.45	123.38	382,450.49	6,412.68
$5,000+	643	728.69	63.07	8,246.94	106.58	468,549.78	4,615.05
C1							
No earnings	2,225	2,531.81	88.37	26,252.60	81.13	5,633,284.59	12,804.37
$1-$5,000	363	1,025.62	97.72	9,535.06	122.52	372,299.98	7,317.34
$5,000+	524	856.85	120.00	16,219.18	135.82	448,991.44	6,681.13

Exhibit B-14. Summary of the Weights After Trimming

Domain	Cases (n)	Mean	Min	Max	CV (%)	Sum
T1						
No earning	2,163	2,590.65	40.88	15,833.20	94.12	5,603,576
$1-$5,000	428	893.58	86.96	6,468.45	119.50	382,450
$5,000+	643	728.69	64.10	4,690.57	97.04	468,550
C1						
No earning	2,225	2,531.81	88.74	12,858.69	78.37	5,633,285
$1-$5,000	363	1,025.62	98.75	7,394.35	117.55	372,300
$5,000+	524	856.85	123.64	6,884.19	112.12	448,991

Component 7: Re-raking

We repeated the raking adjustment discussed in Component 5, applied to the Component 6 weights. A second iteration of raking serves as a smoothing adjustment to recover the lost weight from trimming. In addition, the weights are ratio adjusted so that the sums for respondents plus ineligible cases in T1 and C1 are the same for ten 2011 baseline characteristics and three 2014 outcomes shown in Exhibit B-11. At the end of the weighting process, we saved the re-raked weights as the final analysis weights, **S1_SURVEYWGT.** These are the weights we used when analyzing Stage 1 36-Month Survey data for purposes of generalizing findings to the national BOND population. The sum of the weights for respondents only is an estimate of the number of beneficiaries in the BOND population who were eligible for the survey. Statistics for the final weights appear in Exhibit B-15.

Exhibit B-15. Summary of the Final Weights after Re-raking

Sample Group	Cases (n)	Mean	Min	Max	CV (%)	Sum
Respondents and ineligible cases						
T1	3,234	1,995.85	39.69	18,115.05	111.64	6,454,576
C1	3,112	2,074.09	86.28	13,561.57	93.39	6,454,576
Respondents only						
T1	2,916	2031.39	62.764	18115.05	112.43	5,923,536
C1	2,819	2087.95	86.283	13561.57	95.07	5,885,939

B.3.4. Analysis Modification for Respondents from Multi-subject households.

As discussed in Appendix A, we dropped cases from that sample that were originally selected during Stage 1 sampling but later identified as contaminated because beneficiaries in the same household had been assigned to different demonstration groups; the administrative weights are adjusted for these exclusions. These adjustments are needed to eliminate possible bias in impact estimates based on outcomes in the administrative data for the full Stage 1 sample, because the percentage of subjects from such multi-person households (MPHH) remaining in the T1 sample—subjects that represent the entire population of MPHH beneficiaries—is a fraction of the percentage remaining in the C1 sample, even though the latter is quite small. As result, the weights for MPHH subjects in the T1 sample are much higher than for otherwise comparable subjects (based on administrative characteristics) in the T1 sample, whereas weights for those in the C1 sample are only marginally higher than for otherwise comparable subjects in the C1 sample.

The weights for the survey respondents reflect the administrative weights, because the latter are used in construction of the former. In fact, only 3 MPHH subjects are included among the T1 respondents, compared to 84 among the C1 respondents, reflecting the relatively small number of such subjects in the full administrative sample for T1. In the process of constructing the weights for survey respondents, these 3 subjects had exceptionally high weights before the final trimming and raking steps, and their weights were trimmed as a result. Although the final raking step serves to control for observable characteristics of these subjects, it does not control for any observed characteristics that make them different than observationally equivalent subjects in other households. Furthermore, a regression analysis showed that the MPHH indicator is a statistically significant predictor of treatment versus control membership status (at a p-value of 0.008) using the final survey weights. Hence, there is some potential for bias in the impact estimates based on weighted survey data. Although any such bias is likely to be small, because MPHH cases constitute a small share of the sample even after weighting, to mitigate any possible bias we included an MPHH status indicator as a control variable in the regression models for impact estimation.

Appendix C. Methodology for Overpayments Estimation

In this appendix, we build on the information presented in Section 2.2.2 to describe in more detail our approach to estimating work-related overpayments and incorrect payments. We lay out our methodology in four steps and then discuss measurement validation and analysis limitations.

SSA may overpay beneficiaries for numerous reasons. To focus on the overpayment outcome most likely to be affected by the benefit offset rules and BOND implementation we identify only work-related overpayments and incorrect payments that occur during the BPP (for treatment subjects) and EPE (for control subjects). The primary distinction between overpayments and incorrect payments is the manner in which SSA reconciles the improper payment (see Section 8.1 for more details). We refer to these collectively as "overpayments."

After treatment and control group beneficiaries have exhausted their TWP and GP months, the benefit rules of the SSDI program and the BOND demonstration call for benefit reduction (T1 subjects), suspension (C1 subjects), or termination (C1 subjects only during the time period of analysis) in months in which substantial work activity occurs. These adjustments are not always made on the most expeditious basis. Program rules require beneficiaries to report their earnings to SSA promptly; however, some do not do so. Furthermore, SSA also has a considerable backlog in its processing of work CDRs. Because of these and other reasons, there is often a delay from the first month in which benefits should be adjusted and the month when the adjustment occurs. A delay in benefit adjustment can result in overpayment of the beneficiary's SSDI benefit.

In addition, some situations that uniquely apply to treatment group subjects may also yield overpayments. Overpayments can occur for treatment subjects who submit inaccurate or revised AEEs and also as a result of BSAS errors. AEEs and BSAS are not relevant to control subjects and therefore cannot yield overpayments for those beneficiaries.

As SSA learns of and processes information on past work activity, SSA makes retroactive adjustments to benefits and also updates information in its administrative data system. Historical Disabled Beneficiary and Dependent (DBAD) file extracts provide a record of SSA's knowledge of the beneficiary's work activity and benefits due at that time, which we refer to as "contemporaneous" data. Recent DBAD extracts provide "updated" information about SSA's knowledge of past work activity.[140] For example, if in January 2015 SSA determined that a beneficiary worked during the EPE between May 2011 and December 2011, DBAD extracts from January 2015 on will indicate benefit suspension from May 2011 through December 2011. DBAD extracts before 2015 will indicate that the beneficiary was due a benefit check for those months. Both extracts will indicate that SSA actually paid a benefit in each of those same months.

[140] SSA does not update the benefits due variable in the DBAD for beneficiaries subject to current law or for treatment subjects in full offset (treatment group subjects whose earnings are high enough that their benefit is offset to $0). SSA only adjusts the DBAD benefits due variable for treatment beneficiaries in partial offset (treatment group subjects who receive partial benefits under the $1 for $2 offset formula).

C.1 Steps to Estimate Overpayments

Here we describe how we exploit the difference in monthly DBAD extracts to identify overpayments. The basic computation of an overpayment under this approach is the difference between contemporaneous benefits due (the amount determined to be due for each month *in that month*) and updated benefits due (a more recent value that reflects retroactive adjustments). An alternative is to construct overpayments based on differences between benefits paid and updated benefits due. However, benefits paid may reflect SSA adjustments not related to work activity. For example, SSA lump-sum transfers to reconcile past underpayments are reflected in benefits paid and a comparison of benefits paid to updated benefits due (including the lump-sum transfer) would incorrectly flag this scenario as an overpayment. The difference between updated and contemporaneous benefits due more accurately captures overpayments.

Step 1: Identify Disabled-Worker Beneficiaries

We begin our analysis of overpayments by identifying disabled-worker beneficiaries in the BOND Stage 1 analysis sample who are entitled to SSDI on the basis of their own earnings histories only. We focused on these beneficiaries to avoid potential complications to our method associated with dually-entitled and auxiliary beneficiaries. Specifically, it is difficult in such cases to distinguish between benefit changes due to the primary beneficiary's earnings from those due to own earnings. In addition, dually-entitled beneficiaries may be entitled to multiple TWPs, thus making it very difficult to establish when beneficiaries were working beyond the TWP and GP and hence potentially overpaid. SSA may overpay auxiliary beneficiaries based on work activity of the primary beneficiary.

We identify disabled-worker beneficiaries based on account information in the DBAD. We retain records where the beneficiary is entitled only to his or her own SSDI claim account, on the basis of own earnings. In cases where multiple beneficiaries are entitled to SSDI benefits on a single claim account, we retain only the beneficiary whose earnings are the basis for entitlement. For inclusion in our analysis sample, in each calendar year we require beneficiaries to be disabled-worker beneficiaries and have DBAD records in all months of that calendar year, starting with the first month in which offset use was possible: May 2011.[141] Consistent with all impact analyses, we also exclude beneficiaries in families in which two or more disabled beneficiaries are assigned to different experimental groups. After these exclusions, we retain 82 percent of T1 subjects and 80 percent of C1 subjects. The difference between these two percentages likely reflects differences in the characteristics of beneficiaries in the two groups that are a known consequence of the sample selection process.[142] The final analysis sample sizes include approximately 65,000 T1 subjects and nearly 720,000 C1 subjects (Exhibit C-1).

---

[141] We also apply our algorithm to Stage 2 subjects and will present results in a future Stage 2 report. The first Stage 2 subjects enrolled in March 2011 and were eligible to use the offset in April 2011. Statistics for T1 subjects start in May 2011 because they were all assigned in April 2011 and only potentially eligible to use the offset in May 2011.

[142] Specifically, as described in the *Evaluation Analysis Plan*, compared to C1, T1 includes a larger share of short-duration beneficiaries and a smaller share of concurrent beneficiaries. The impact analysis uses weights to correct for these differences. The *Stage 1 Snapshot Report* finds that, at baseline, 88.5 percent of T1 subjects and 88.8 percent of C1 subjects were primary beneficiaries (i.e. disabled-worker beneficiaries) with or without entitlement to benefits from another source (e.g. auxiliary entitlement). In this analysis, we require that a beneficiary be a disabled-worked beneficiary *only* (e.g. no auxiliary entitlement) and continuously over the course of nine months to a year, rather than at a point in time.

Exhibit C-1. Sample Sizes for Overpayment Analysis

	2011	2012	2013
T1	65,234	65,337	65,338
C1	718,839	719,301	719,289

Step 2: Identify Beneficiaries Eligible for an Overpayment

The first step in calculating overpayments is to identify beneficiaries eligible for such overpayments. Eligibility varies by BOND treatment group because of the difference in BOND treatment rules relative to current law. We use updated data from the October 2015 DBAD to flag beneficiary-months in 2011, 2012, and 2013 during which overpayments are possible. Specifically, we identify months as eligible for an overpayment when they meet any of the following criteria:

- For the C1 group, months with earnings above SGA after the GP during or after the EPE. In accordance with current-law rules, SSA should suspend or terminate benefits in these months.
- For the T1 group, months with earnings above BYA after the GP during the BPP. In accordance with BOND rules, SSA should reduce or suspend benefits in these months—these are beneficiaries for whom SSA should adjust benefits under the offset.

In other words, we identify beneficiaries who had (or should have had) their benefits reduced, suspended, or terminated as a result of work activity, as the only beneficiaries who could have incurred an overpayment. If SSA did not make the necessary adjustment in the first month for which such an adjustment was warranted, overpayments likely occurred monthly from that point until either SSA made the adjustment or the beneficiary reduced her or his earnings.[143]

Step 3: Determine if the Beneficiary was Overpaid

Next, we analyze the records of beneficiaries identified in Step 2 to determine whether or not the beneficiary was overpaid. Because of the difference in the way work activity affects benefits, our method for identifying overpayments differs for treatment and control group subjects.

Benefit receipt is binary for control group subjects: the beneficiary was either entitled to a full check on the basis of own work activity or was not so entitled. We classify control group beneficiaries as overpaid during months in 2011, 2012, and 2013 in which they worked during or after the EPE according to updated data in the October 2015 DBAD extract and, according to the same source, reflect that the beneficiary received a benefit check. Because these beneficiaries were not entitled to a benefit check, the benefit due was an overpayment.

[143] Note that we do not attempt to identify overpayments accrued during the initial reinstatement period (IRP) after expedited reinstatement (EXR). Beneficiaries who terminate from the SSDI program for work may subsequently become re-entitled through EXR. Upon re-entitlement, the beneficiary enters the IRP and must complete 24 months without engaging in SGA before returning to standard SSDI or BOND work rules.

Treatment group subjects who work during the BPP may receive partial benefits under the $1 for $2 offset (referred to as "partial offset") unless their earnings are so high that they are not entitled to receive any DI cash benefit (referred to as "full offset"). We apply the same logic used to identify overpayments among control group subjects to treatment group subjects in full offset: we classify treatment group beneficiaries as overpaid during months in 2011, 2012, and 2013 in which they are in full offset and received a benefit check, according to the updated DBAD extract. For treatment group subjects in partial offset, we compare data from months in 2011, 2012, and 2013 in the updated October 2015 DBAD data to the contemporaneous data in the monthly DBAD files from 2011, 2012, and 2013. We flag as an overpayment any month in which (1) the beneficiary was in partial offset according to the updated data, (2) the updated data indicate that SSA paid a benefit, and (3) the contemporaneous benefit due was greater than the updated benefit due.

Step 4: Calculate the Overpayment Amount

We calculate the overpayment amount for beneficiary-months flagged with an overpayment as follows.

For control group beneficiaries, the overpayment amount is equal to the monthly benefit due variable in that month based on the updated data. Although conceptually the monthly benefit due in the updated data should be reduced to $0, SSA does not adjust the administrative benefit due variable in the updated data for beneficiaries whose benefits are entirely suspended (C1 subjects and T1 subjects in full offset). Because SSA does not adjust the updated data to reflect that the beneficiary was due $0, the positive benefit due amount in the updated data represents the overpayment amount. SSA does adjust the administrative data to update benefits due for BOND treatment subjects in partial offset.

For treatment group beneficiaries, the overpayment amount is measured as the difference between the contemporaneous benefit due and either (a) the updated benefit due for beneficiaries in partial offset or (b) $0 for beneficiaries in full offset.

C.2 Validation

We validated the results of this measurement procedure in several ways. First, we compared DBAD statistics on the number of T1 subjects in the offset against monthly statistics the Implementation Team publishes on the number of identified offset users. As of the end of October 2015, Implementation Team records identified 1,857 T1 subjects who used the offset by December 2013. We identified the same number of T1 subjects in the October 2015 DBAD. Reflecting the exclusion criteria used to establish the overpayment analysis sample (auxiliary beneficiaries and disabled workers who are dually entitled or who have incomplete DBAD records), 1,646 of these T1 subjects are in the overpayment analysis sample.

As an additional validation check, we used the 2013 Disability Analysis File (DAF) to construct a measure of overpayments using a different methodology. The underlying administrative data used to create the 2013 DAF results were extracted on March 28, 2014. To make the timing consistent, we compared results using the DBAD algorithm based on the April 2014 DBAD. Beyond timing, there are additional differences between the DBAD and DAF data sources including differences in data available to identify disabled-worker beneficiaries and differences in the historical marker of benefits (benefits due versus benefits paid). Specifically, the DAF calculation is based on differences between benefits paid and

updated benefits due.[144] Because of these differences, our algorithms differ based on data source. We believe the DBAD method more accurately captures the experiences of BOND beneficiaries.

There is broad consistency in aggregate across the data sources (Exhibit C-2). Unconditional overpayment prevalence rates are within 7 to 20 percent of each other and conditional rates are within 1 to 5 percent. This suggests that the DAF and DBAD identify beneficiaries with substantial employment after the GP at different rates but among those subgroups eligible for an overpayment, the algorithms identify beneficiaries with such overpayments at similar rates. The DBAD and DAF algorithms identify unconditional overpayments within 2 to 20 percent of each other and conditional overpayments within 8 to 13 percent of each other. Relative to the DAF, the DBAD produced higher estimates of overpayment amounts for T1 subjects and the reverse is true for C1 subjects. A priori, there was no expectation about the relative size of the DBAD and DAF estimates for the proportion overpaid or the size of the overpayment. This is because the DAF benefits paid variable may include payment transfers that are both higher and lower than the DBAD benefits due variable.

Exhibit C-2. DBAD and DAF Statistics on Overpayments

Outcome	T1 Mean		C1 Mean	
	DBAD	DAF	DBAD	DAF
Percentage with Overpayment				
All subjects overpaid in any month in 2011 or 2012 (%)	1.13	1.06	1.36	1.13
Subjects eligible for an overpayment overpaid in any month in 2011 or 2012 (%)	71.8	68.6	63.5	64.2
Mean Overpayment Amount				
Mean combined 2011 and 2012 overpayment among all subjects	$55	$56	$118	$94
Mean combined 2011 and 2012 overpayment among overpaid subjects	$4,846	$5,247	$8,673	$7,530

Source: 2013 DAF and DBAD extracts from May 2011-December 2012, and April 2014. Dollar amounts are not adjusted for inflation.

The DBAD estimates also seem reasonable compared to a related statistic recently published by the SSA Office of the Inspector General (OIG, 2015). The OIG found that 32 of 985 DI beneficiaries (3.2 percent) in a representative sample on the SSDI rolls in 2003 had overpayments for reasons related to work activity or income over the 11 years (2003 through 2013).[145] We expect the rate of overpayments among

[144] The DAF measure of benefits paid reflects all adjustments to benefits and is more volatile than the DBAD measure of benefits due. While we attempt to make minor adjustments to account for such transfers in the DAF algorithm, it is likely that the DAF method may incorrectly classify some work-related overpayments and incorrect payments.

[145] The OIG sample randomly selected 1,532 SSA disability beneficiaries for their study (including 547 SSI-only beneficiaries) who were ages 18 to 61, within two strata based on diagnosis codes using proportional allocation methodology. The report indicated that the sample represented the population. The age criteria are similar to that imposed in BOND, which includes beneficiaries ages 20 to 59 in 2011 and 22 to 61 in 2013.

current-law beneficiaries to be the same as among C1 subjects, who are representative of the broader SSDI population and also subject to current-law rules. If we assume the same rate of new C1 overpayment accruals that occurred between the second and third years of the demonstration continues for seven more years, the percentage of C1 subjects accruing an overpayment over 10 years would be 4.8 percent, within the vicinity of the OIG's finding of 3.2 percent over 11 years.[146] However, among a cohort, we expect the rate of overpayment accrual to continue to slow over time because beneficiaries are less likely to work (and thus less likely to be at risk of an overpayment) with longer durations on the SSDI rolls (Liu and Stapleton 2010). Hence, the DBAD estimate appears in line with the SSA Office of the Inspector General estimate.

Our final approach to validation was via SSA case reviews. A member of the ORDES work unit conducted in-depth case reviews of 60 randomly selected records to construct an SSA measure of overpayments to compare to our algorithm results. We randomly selected an equal number of treatment and control group cases with a mix of overpayments and no overpayments as follows: 10 from each group with no overpayment, and 20 from each group with an overpayment. That is, there are four groups of cases.

Among cases for which the DBAD algorithm suggested there were no overpayments—the first two groups of cases, the SSA calculations are generally in agreement. Both sets of results indicated no overpayments for 80 and 90 percent of treatment and control cases reviewed, respectively (Exhibit C-3).

Among the 10 control group cases with no overpayment according to the DBAD algorithm, there is one for which the SSA method identifies an overpayment. That case had a $2 overpayment according to the SSA calculations. Although there is a discrepancy between the two methods, the SSA-identified overpayment amount falls within an established current-law standard for determining whether an overpayment is large enough to warrant action. Under current law, SSA allows for certain administrative tolerances when determining whether or not to officially record the overpayment on the beneficiary's record and notify the beneficiary of the overpayment. In cases in which the beneficiary is alive, has a manually computed overpayment less than $30, and SSA is not preparing a notice for a reason other than the overpayment, SSA will not pursue further action (Program Operations Manual System GN 02201.013).[147] Although this standard was established for administrative convenience, rather than accuracy, it appears to be a reasonable guideline for determining an absolute tolerance level for the purposes of comparing outcomes from the two overpayment sources.

The SSA method identifies overpayments for 2 of the 10 treatment cases for which the DBAD algorithm indicates no overpayment. Both cases reflect conceptual differences between the SSA method and the DBAD algorithm. The SSA method shows an overpayment for one case caused by a technician error. In

[146] The DBAD estimates diminishing numbers of new C1 subjects with overpayments each year: an increase of 0.46 percentage points between 2011 and 2012 and an increase 0.42 percentage points between 2012 and 2013. We computed the 10-year accrual rate as the proportion with overpayments in the first three years as determined by the DBAD plus the 0.42 percentage point new accrual rate for the seven remaining years: 1.85 + 7*0.42 = 4.8.

[147] Under BOND rules, SSA does not notify or seek to reconcile overpayments less than $200 (Stapleton et al. 2010).

this case, a SSA technician issued a check for nearly $24,000 in November 2013 and the SSA method identifies that amount as an overpayment. Although this overpayment occurred while the beneficiary was working, the overpayment is not related to the beneficiary's earnings. Hence, despite the mismatch between the two sources, this does not indicate a flaw in the DBAD's logic. In the second case, the beneficiary was in EPE suspense before BOND and did not receive any checks in 2012. The beneficiary was entitled to a partial benefit payment in that year as a result of BOND and, in February 2014, SSA issued the beneficiary a lump sum payment to reconcile the underpayment. However, the lump-sum check covered payments SSA owed outside the time period of analysis and the SSA method classifies this lump sum as an overpayment because it was greater than the actual benefits due during the period of analysis. Although there is a mismatch between the two sources, this also does not indicate an error in the classification derived under the DBAD method.

Exhibit C-3. Comparison of DBAD Statistics and SSA Case Reviews: Records with No Overpayment According to the DBAD Algorithm

	Total Cases	DBAD Estimate and SSA Calculation Match Exactly	DBAD Estimate Within $0.01 and $30 of SSA calculation	DBAD Estimate More than $30 Different than SSA Calculation
Treatment	10	8	0	2
Control	10	9	1	0

Sources: SSA Single Copy estimations produced by SSA and authors' calculations based on DBAD extracts from May 2011-December 2012, and April 2014.

All cases for which the DBAD indicated overpayments—the third and fourth groups of cases—also had overpayments according to the SSA calculations, although the size of the overpayment varied across the two sources. The DBAD estimates either match or are within $30 or 5 percent of the SSA overpayment (or both) calculations for 9 of the 20 reviewed treatment cases and 16 of the 20 control cases (Exhibit C-4).[148] Half of treatment cases and 3 control cases were greater than 5 percent and less than 20 percent of the SSA calculations. These differences generally occur because of the DBAD algorithm does not account for portions of the overpayment related to payment changes such as lump sum adjustments for Medicare premium payments,[149] changes due to primary insurance amount (PIA) recomputations,[150] benefit garnishment,[151] and recovery of pre-BOND overpayments. It is important to know that the mismatches arise from situations that may occur for both treatment and control group cases.

[148] The SSA single copy method only reflects overpayments that have yet to be recovered, as of the date of analysis. Three reviewed cases (one treatment and two control cases) had their overpayments recovered and did not show overpayments on the single copy. However, ignoring the repayment, the overpayment amount accrued during the analysis period reflect in the DBAD matches the adjusted SSA single copy results.

[149] Medicare Part B (supplemental medical insurance) and Part D (prescription drug coverage) premiums are typically taken directly from the SSDI cash benefit. When benefits are suspended due to work, beneficiaries must pay these Medicare premiums or will be disenrolled from coverage.

[150] PIA recomputations occur when SSA retroactively adjustments monthly benefit amounts based on changes in earnings, most frequently as part of the Automatic Earnings Reappraisal Operations (AERO) process.

[151] SSDI benefits may be garnished to enforce child support or alimony obligations and to collect overdue federal taxes, among other reasons.

The DBAD overpayment estimates for the remaining cases—one each in the control and treatment groups—were both more than 20 percent different than the SSA estimates. The DBAD algorithm estimated a $959 overpayment for the treatment case relative to a $184 overpayment reported through the SSA case review, a difference of $775. The difference appears to be due to multiple changes in Medicare premium payments and premium refunds. For the control case, the DBAD algorithm estimated a $1,386 overpayment relative to a $1,865 overpayment reported through the SSA case review, a difference of $479. This is due to a PIA recomputation that resulted in a *lower* PIA amount and did not enter into the DBAD algorithm's metrics. According to an ORDES BOND work unit staff person, this is an extremely rare occurrence.

Exhibit C-4. Comparison of DBAD Statistics and SSA Case Reviews: Records with Overpayments According to the DBAD Algorithm

	Total Cases	DBAD Estimate and SSA Calculation Match Exactly	DBAD Estimate Within $0.01 and $30 or 0.1% and 5% of SSA Calculation	DBAD Estimate Within 5.1% and 10% of SSA Calculation	DBAD Estimate Within 10.1% and 20% of SSA Calculation	DBAD Estimate Over 20% of SSA Calculation
Treatment	20	7	2	2	8	1
Control	20	9	7	3	0	1

Sources: SSA Single Copy estimations produced by SSA and authors' calculations based on DBAD extracts from May 2011-December 2012, and April 2014.

In aggregate, there is consistency between the two methods in the size of the overpayments. Among the 20 treatment group cases reviewed for which the DBAD algorithm indicated overpayments, on average the DBAD estimate is 3.9 percent lower than the SSA estimate. The 20 control group cases with overpayments based on the DBAD algorithm is 0.3 percent smaller than the corresponding SSA estimate. Hence, despite more substantial differences in some individual cases, the DBAD algorithm produces results that, in average, agree closely with the SSA calculations. The number of cases reviewed is too small to infer that these findings extend to all estimates based on the DBAD algorithm.

C.3 Limitations

The DBAD algorithm has several known limitations. First, the DBAD algorithm does not encompass all overpayment scenarios that can occur for DI beneficiaries. In some non-standard situations not captured in the DBAD data, such as when SSA issues beneficiaries emergency checks in a field office or erroneous checks due to technician error, the algorithm may not accurately predict the presence or size of an overpayment. The DBAD algorithm will also not capture payment changes related to Medicare payments AERO adjustments, benefit garnishment, and overpayment recovery. However, the DBAD may produce estimates that are higher or lower than SSA records such that we expect the aggregate statistics to be a close approximation.

Finally, the DBAD algorithm does not identify work-related underpayments. Our approach cannot be used to identify of work-related underpayments because, unlike work-related overpayments, they can occur when the beneficiary is not working. For example, if a beneficiary lost his job in March and SSA

didn't increase his benefits until July, the beneficiary would have been underpaid for March, April, and June, but he would not be flagged as a worker in those months. In contrast, for work-related overpayments, we are able to use evidence of earnings above BYA (treatment subjects) or of engagement in SGA during the EPE (control subjects) to distinguish between work-related overpayments and other overpayments.

Appendix D. Subgroup Exhibits for 2014 Earnings and Benefit Impacts

Exhibit D-1. Estimated Impacts on 2014 Outcomes of the Stage 1 Intervention for Subgroups Defined by Duration of SSDI Receipt

Outcome	Short Duration			Long Duration			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2014)							
Total earnings	$1,626	$1,635	-$9 ($40)	$1,281	$1,253	$29 ($36)	-$37 ($54)
Employment during year (%)	14.73	14.48	0.25 (0.21)	12.52	12.22	0.30 (0.17)	-0.05 (0.27)
Earnings above BYA (%)	3.45	3.31	0.15 (0.11)	2.65	2.43	0.22* (0.10)	-0.08 (0.15)
Earnings above 2 times BYA (%)	1.45	1.60	-0.15** (0.07)	1.14	1.14	-0.01 (0.06)	-0.15† (0.09)
Earnings above 3 times BYA (%)	0.85	0.92	-0.07 (0.05)	0.55	0.62	-0.07 (0.04)	0.00 (0.07)
Benefit Outcomes (January–December 2014)							
Total SSDI benefits paid	$11,987	$11,915	$71** ($31)	$10,903	$10,751	$152*** ($36)	-$81* ($48)
Number of months with SSDI payments	10.45	10.38	0.08** (0.03)	10.52	10.40	0.11*** (0.02)	-0.04 (0.04)
Total SSI benefits paid	$350	$358	-$7 ($8)	$446	$450	-$4 ($10)	-$3 ($13)
Number of months with SSI payments	1.40	1.44	-0.04 (0.02)	1.98	1.99	-0.01 (0.02)	-0.03 (0.03)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 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. Dollar values are not adjusted for inflation.

Unweighted sample sizes: Short duration T1 = 38,669, Short duration C1 = 209,790, Long duration T1 = 38,446, Long duration C2 = 681,808

*/**/*** 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 (and with no multiple-comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit D-2. Estimated Impacts on 2014 Outcomes of the Stage 1 Intervention for Subgroups Defined by Baseline SSI Status

Outcome	SSDI-Only			Concurrent			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2014)							
Total earnings	$1,520	$1,491	$29 ($36)	$780	$813	-$33 ($37)	$61 ($51)
Employment during year (%)	13.49	13.15	0.34** (0.15)	11.81	11.80	0.02 (0.28)	0.32 (0.31)
Earnings above BYA (%)	3.16	2.92	0.24** (0.08)	1.67	1.66	0.01 (0.20)	0.23 (0.22)
Earnings above 2 times BYA (%)	1.43	1.46	-0.03 (0.05)	0.33	0.46	-0.13 (0.07)	0.10 (0.09)
Earnings above 3 times BYA (%)	0.76	0.84	-0.08* (0.04)	0.13	0.15	-0.02 (0.04)	-0.06 (0.05)
Benefit Outcomes (January–December 2014)							
Total SSDI benefits paid	$12,454	$12,316	$138*** ($29)	$5,710	$5,630	$80 ($52)	$58 ($59)
Number of months with SSDI payments	10.57	10.46	0.10*** (0.02)	10.18	10.08	0.10 (0.06)	0.01 (0.06)
Total SSI benefits paid	$38	$38	-$1 ($2)	$2,127	$2,153	-$26 ($40)	$25 ($40)
Number of months with SSI payments	0.18	0.17	0.00 (<0.01)	9.15	9.26	-0.10 (0.07)	0.11 (0.07)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 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. Dollar values are not adjusted for inflation.

Unweighted sample sizes: SSDI-only T1 = 64,709, SSDI-only C1 = 694,270, Concurrent T1 = 12,406, Concurrent C1 = 197,328

*/**/*** 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 (and with no multiple-comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit D-3. Estimated Impacts on 2014 Outcomes of the Stage 1 Intervention for Subgroups Defined by Employment in 2010

Outcome	Employed			Not Employed			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2014)							
Total earnings	$5,551	$5,617	-$66 ($166)	$526	$492	$34 ($23)	-$100 ($167)
Employment during year (%)	48.41	47.63	0.78 (0.57)	5.92	5.73	0.19 (0.11)	0.59 (0.58)
Earnings above BYA (%)	11.37	10.80	0.57 (0.39)	1.14	1.02	0.12** (0.05)	0.44 (0.40)
Earnings above 2 times BYA (%)	5.02	5.45	-0.42 (0.25)	0.45	0.42	0.02 (0.03)	-0.45† (0.25)
Earnings above 3 times BYA (%)	2.73	3.14	-0.42* (0.19)	0.21	0.21	0.00 (0.03)	-0.41†† (0.20)
Benefit Outcomes (January–December 2014)							
Total SSDI benefits paid	$12,034	$11,592	$442*** ($67)	$11,064	$10,998	$66** ($28)	$376††† ($73)
Number of months with SSDI payments	10.61	10.22	0.39*** (0.07)	10.47	10.43	0.05** (0.02)	0.34††† (0.07)
Total SSI benefits paid	$208	$218	-$10 ($9)	$460	$464	-$4 ($9)	-$6 ($13)
Number of months with SSI payments	1.03	1.02	0.00 (0.03)	1.97	1.99	-0.02 (0.02)	0.02 (0.03)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 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. Dollar values are not adjusted for inflation.

Unweighted sample sizes: Employed T1 = 14,694, Employed C1 = 138,194, Not employed T1 = 62,421, Not employed C1 = 753,404

*/**/*** 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 (and with no multiple-comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit D-4. Estimated Impacts on 2014 Outcomes of the Stage 1 Intervention for Subgroups Defined by Access to Medicaid Buy-In Programs

Outcome	Access to Medicaid Buy-In Programs			No Access to Medicaid Buy-In Programs			Estimated Difference in Impact
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	(7)
Earnings and Employment Outcomes (January–December 2014)							
Total earnings	$1,488	$1,462	$26 ($34)	$1,137	$1,141	-$4 ($60)	$30 ($69)
Employment during year (%)	14.21	13.83	0.39** (0.17)	10.70	10.67	0.02 (0.28)	0.36 (0.33)
Earnings above BYA (%)	3.08	2.84	0.24** (0.08)	2.42	2.33	0.09 (0.19)	0.15 (0.20)
Earnings above 2 times BYA (%)	1.34	1.37	-0.03 (0.06)	0.97	1.06	-0.09 (0.07)	0.06 (0.09)
Earnings above 3 times BYA (%)	0.69	0.77	-0.08* (0.04)	0.53	0.57	-0.04 (0.04)	-0.04 (0.06)
Benefit Outcomes (January–December 2014)							
Total SSDI benefits paid	$11,289	$11,164	$125** ($40)	$11,087	$10,951	$136*** ($34)	-$11 ($52)
Number of months with SSDI payments	10.49	10.39	0.11*** (0.02)	10.50	10.41	0.09** (0.03)	0.02 (0.04)
Total SSI benefits paid	$431	$437	-$6 ($10)	$383	$386	-$3 ($10)	-$3 ($14)
Number of months with SSI payments	1.82	1.85	-0.03 (0.02)	1.77	1.76	0.01 (0.02)	-0.04 (0.03)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 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. Dollar values are not adjusted for inflation.

Unweighted sample sizes: Access to Medicaid Buy-In Programs T1 = 48,941, Access to Medicaid Buy-In Programs C1 = 567,760, No access to Medicaid Buy-In Programs T1 = 28,174, No access to Medicaid Buy-In Programs C1 = 323,838

*/**/*** 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 (and with no multiple-comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit D-5. Estimated Impacts on 2014 Outcomes of the Stage 1 Intervention for Subgroups Defined by Age at Baseline

Outcome	Age 49 or Less at Baseline			Age 50 or More at Baseline			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2014)							
Total earnings	$1,977	$1,984	-$7 ($42)	$855	$816	$39 ($37)	-$46 ($56)
Employment during year (%)	17.91	17.62	0.29 (0.22)	8.95	8.67	0.27* (0.15)	0.01 (0.26)
Earnings above BYA (%)	4.38	4.15	0.23 (0.13)	1.56	1.39	0.17** (0.07)	0.06 (0.14)
Earnings above 2 times BYA (%)	1.86	1.92	-0.06 (0.08)	0.66	0.71	-0.05 (0.05)	-0.01 (0.09)
Earnings above 3 times BYA (%)	0.90	1.03	-0.13** (0.05)	0.41	0.43	-0.02 (0.04)	-0.11 (0.07)
Benefit Outcomes (January–December 2014)							
Total SSDI benefits paid	$9,766	$9,589	$177*** ($43)	$12,544	$12,459	$84** ($29)	$92† ($52)
Number of months with SSDI payments	10.36	10.20	0.15*** (0.03)	10.62	10.57	0.06** (0.02)	0.10†† (0.04)
Total SSI benefits paid	$607	$608	$0 ($15)	$246	$256	-$10 ($5)	$9 ($16)
Number of months with SSI payments	2.49	2.52	-0.02 (0.02)	1.19	1.20	-0.01 (0.02)	-0.01 (0.03)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 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. Dollar values are not adjusted for inflation.

Unweighted sample sizes: Age 49 or less at baseline T1 = 36,283, Age 49 or less at baseline C1 = 428,043, Age 50 or more at baseline T1 = 40,832, Age 50 or more at baseline C1 = 463,555

*/**/*** 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 (and with no multiple-comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit D-6. Estimated Impacts on 2014 Outcomes of the Stage 1 Intervention for Subgroups Defined by Primary Impairment of Major Affective Disorder

Outcome	Primary Impairment of Major Affective Disorder			All Other Primary Impairments			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2014)							
Total earnings	$1,556	$1,502	$54 ($65)	$1,354	$1,343	$10 ($29)	$43 ($71)
Employment during year (%)	15.37	15.13	0.24 (0.38)	12.78	12.49	0.29* (0.14)	-0.05 (0.40)
Earnings above BYA (%)	3.53	3.09	0.44* (0.23)	2.77	2.62	0.15* (0.07)	0.29 (0.24)
Earnings above 2 times BYA (%)	1.30	1.36	-0.06 (0.17)	1.22	1.27	-0.05 (0.05)	0.00 (0.17)
Earnings above 3 times BYA (%)	0.61	0.70	-0.09 (0.08)	0.65	0.72	-0.07* (0.04)	-0.02 (0.09)
Benefit Outcomes (January–December 2014)							
Total SSDI benefits paid	$11,232	$11,112	$120* ($61)	$11,230	$11,100	$129*** ($25)	-$9 ($66)
Number of months with SSDI payments	10.88	10.78	0.10* (0.05)	10.43	10.32	0.10*** (0.02)	0.00 (0.06)
Total SSI benefits paid	$449	$446	$3 ($14)	$411	$418	-$7 ($9)	$10 ($17)
Number of months with SSI payments	2.02	2.04	-0.02 (0.04)	1.77	1.78	-0.02 (0.01)	-0.01 (0.05)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 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. Dollar values are not adjusted for inflation.

Unweighted sample sizes: Primary impairment of major affective disorder T1 = 12,024, Primary impairment of major affective disorder C1 = 145,893, All other primary impairments T1 = 65,091, All other primary impairments C1 = 745,705

*/**/*** 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 (and with no multiple-comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit D-7. Estimated Impacts on 2014 Outcomes of the Stage 1 Intervention for Subgroups Defined by Primary Impairment of Back Disorder

Outcome	Primary Impairment of Back Disorder			All Other Primary Impairments			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings and Employment Outcomes (January–December 2014)							
Total earnings	$1,031	$999	$32 ($56)	$1,443	$1,428	$15 ($33)	$17 ($65)
Employment during year (%)	9.91	9.50	0.41 (0.30)	13.72	13.46	0.26 (0.15)	0.15 (0.33)
Earnings above BYA (%)	2.17	1.97	0.20 (0.15)	3.01	2.81	0.20** (0.08)	0.00 (0.17)
Earnings above 2 times BYA (%)	0.91	1.02	-0.11 (0.10)	1.28	1.32	-0.04 (0.05)	-0.07 (0.11)
Earnings above 3 times BYA (%)	0.49	0.58	-0.09 (0.07)	0.67	0.74	-0.07 (0.04)	-0.02 (0.08)
Benefit Outcomes (January–December 2014)							
Total SSDI benefits paid	$13,394	$13,326	$68 ($48)	$10,875	$10,737	$138*** ($30)	-$70 ($56)
Number of months with SSDI payments	11.30	11.21	0.09* (0.04)	10.37	10.26	0.10*** (0.02)	-0.01 (0.05)
Total SSI benefits paid	$191	$186	$6 ($7)	$454	$461	-$7 ($9)	$13 ($11)
Number of months with SSI payments	0.95	0.93	0.02 (0.02)	1.95	1.97	-0.02 (0.02)	0.04† (0.02)

Source: SSA administrative records for calendar year 2014 and baseline SSA administrative data.

Notes: See Chapter 2 for variable definitions. All earnings outcomes are based on a measure of earnings subject to Social Security taxes (see Appendix A.3 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. Dollar values are not adjusted for inflation.

Unweighted sample sizes: Primary impairment of back disorder T1 = 11,698, Primary impairment of back disorder C1 = 116,604, All other primary impairments T1 = 65,417, All other primary impairments C1 = 774,994

*/**/*** 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 (and with no multiple-comparisons adjustment).

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Appendix E. Exhibits for the 36-Month Survey Results

Exhibit E-1. Estimated Differences in Stage 1 Subjects' Understanding of How Earnings Affect Future Benefit Eligibility, by Employment Status in 2010

Outcome	T1			C1		
	Employed in 2010 Mean	Not Employed in 2010 Mean	Difference Estimate	Employed in 2010 Mean	Not Employed in 2010 Mean	Difference Estimate
All Stage 1 Subjects						
Would remain eligible for benefits indefinitely (never have to reapply) (%)	11.6	10.7	0.8 (1.8)	11.7	10.2	1.4 (2.2)
Would remain eligible for a while (eventually would have to reapply) (%)	**60.2**	**58.4**	1.8 (3.1)	**57.4**	**58.2**	-0.8 (3.3)
Would neither remain eligible indefinitely nor for a while (%)	13.2	11.0	2.2 (3.0)	13.4	11.0	2.4 (2.7)
Don't know about future eligibility (%)	15.1	19.9	-4.8* (2.1)	17.6	20.6	-3.0 (2.5)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: 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. Weights reflecting sample selection and survey non-response ensure that the BOND subjects in both samples who met analysis criteria are representative of SSDI recipients in the nation on almost all observed characteristics. There is one discrepancy: the weights 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). Any effect of this discrepancy on the reported statistics is small, however, because the percentage of such subjects in the national BOND-eligible population is less than 3 percent. Standard errors appear in parentheses.

Unweighted sample sizes: Employed in 2010 T1 = 1,106, C1 = 962; not employed in 2010 T1 = 1,180, C1 = 1,857

*/**/*** 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 (with no multiple-comparisons adjustment).

Exhibit E-2. Estimated Impacts on Employment Supports of Stage 1 Subjects Since Random Assignment

Outcome	Treatment Mean	Control Mean	Impact Estimate
Received any type of employment support since random assignment (RA) (%)	53.2	52.5	0.7 (1.9)
Type of Employment Support Received Since RA[a]			
Work or job assessment (%)	2.3	3.2	-0.8 (0.6)
Help to find job (%)	4.8	5.1	-0.2 (0.9)
Training to learn new job or skill (%)	5.1	4.7	0.4 (0.7)
Advice about modifying job or workplace (%)	3.1	3.4	-0.3 (0.7)
Received on-the-job training, coaching, or support services (%)	5.7	6.7	-1.0 (1.0)
Personal care assistance (%)	2.0	2.1	-0.1 (0.4)
Job coach (%)	1.4	1.7	-0.2 (0.5)
Personal care assistance other than job coach (%)	0.6	0.4	0.2 (0.2)
Transportation assistance (%)	28.5	27.4	1.1 (1.7)
Help in keeping a job (%)	2.1	3.0	-0.9* (0.5)
Any kind of assistive device (%)	22.3	22.8	-0.5 (1.6)
Other (%)	2.8	4.1	-1.2 (0.7)

Source: BOND Stage 1 36-Month Survey.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics. All T1 and C1 beneficiaries were asked about their receipt of employment supports.

[a] The sum of means in this section may exceed the mean reported in the same column in the first row because subjects may have received several types of employment support.

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 (with no multiple-comparisons adjustment).

Exhibit E-3. Estimated Impacts on Unmet Need for Employment Support of Stage 1 Subjects Since Random Assignment

Outcome	Treatment Mean	Control Mean	Impact Estimate
Needed any type of employment support but received none since RA (%)	37.3	37.7	-0.3 (2.0)
Type of Employment Support Received Since RA[a]			
Needed but did not receive a work or job assessment (%)	6.8	6.1	0.7 (1.1)
Needed but did not receive help to find job (%)	8.7	9.5	-0.8 (1.2)
Needed but did not receive training to learn new job or skill (%)	9.6	11.1	-1.5 (1.5)
Needed but did not receive advice about modifying job or workplace (%)	4.2	5.0	-0.8 (0.9)
Needed but did not receive on-the-job training, coaching, or support services (%)	7.5	7.2	0.3 (1.2)
Needed but did not receive personal care assistance (%)	5.9	6.6	-0.7 (1.4)
Needed but did not receive transportation assistance (%)	9.7	12.4	-2.6** (1.2)
Needed but did not receive help in keeping a job (%)	6.5	6.5	0.0 (1.0)
Needed but did not receive any kind of assistive device (%)	9.6	10.9	-1.3 (1.3)
Needed but did not receive other employment support (%)	3.1	3.4	-0.3 (0.7)

Source: BOND Stage 1 36-Month Survey.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics. All T1 and C1 beneficiaries were asked about their unmet need for employment supports.

[a] The sum of means in this section may exceed the mean reported in the same column in the first row because subjects may have an unmet need for several types of employment support.

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 (with no multiple-comparisons adjustment).

Exhibit E-4. Estimated Impacts on Education and Training of Stage 1 Subjects Since Random Assignment

Outcome	Treatment Mean	Control Mean	Impact Estimate
Received any schooling or training since RA (%)	6.0	5.6	0.4 (0.8)
Hours per week in training or schooling since RA	1.1	1.3	-0.2 (0.2)
Schooling and Training Types Attended During 3 Years After RA[a]			
Regular high school or directed toward a high school diploma (%)	0.0	0.0	0.0 (0.0)
Preparation for a GED examination (%)	0.3	0.1	0.3 (0.2)
Two-year college directed toward a degree (%)	0.9	0.8	0.1 (0.3)
Four-year college directed toward a degree (%)	0.6	0.4	0.1 (0.3)
Graduate courses (%)	0.1	0.2	-0.1 (0.1)
College courses not directed toward a degree (%)	0.3	0.7	-0.4 (0.3)
Vocational education outside a college (%)	1.4	1.9	-0.5 (0.7)
Non-vocational adult education not directed toward a degree (%)	0.4	0.6	-0.2 (0.2)
Job search assistance, job finding, orientation to the world of work (%)	1.5	1.1	0.4 (0.5)
Other (%)	1.1	0.9	0.2 (0.3)

Source: BOND Stage 1 36-Month Survey.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

[a] The sum of means in this section may exceed the mean reported in the same column in the first row because subjects may have attended several school and training types.

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 (with no multiple-comparisons adjustment).

Exhibit E-5. Estimated Impacts on Education Outcomes for Stage 1 Subjects at the Time of Survey Interview

Outcome	T1 Mean	C1 Mean	Impact Estimate
Currently enrolled in school or taking classes (%)	3.2	3.3	-0.1 (0.8)
Currently working toward a degree, certificate, or license (%)	2.0	2.2	-0.3 (0.5)
Enrolled in school full-time (%)	0.9	1.2	-0.2 (0.3)
Enrolled in school part-time (%)	1.0	1.1	0.0 (0.4)
Currently working toward a GED or high school equivalence program completion (%)	0.3	0.1	0.2 (0.2)
Currently working toward a vocational or training program completion (%)	0.5	0.5	0.0 (0.3)
Currently working toward an associate's degree (%)	0.3	0.5	-0.2 (0.2)
Currently working toward a bachelor's degree (%)	0.4	0.6	-0.2 (0.2)
Currently working toward a graduate degree (%)	0.3	0.3	0.0 (0.2)
Currently working toward other degree or certificate type (%)	0.1	0.1	0.0 (0.1)
Currently not working toward a degree, certificate, or license and only taking classes (%)	1.1	1.1	0.0 (0.4)

Source: BOND Stage 1 36-Month Survey.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

Exhibit E-6. Estimated Impacts on Training to Learn a New Job or Skill for Stage 1 Subjects at the Time of Survey Interview

Outcome	T1 Mean	C1 Mean	Impact Estimate
Received training to learn a new job or skill (%)	5.1	4.7	0.4 (0.7)
Source of Training to Learn a New Job or Skill[a]			
Vocational rehabilitation agency (%)	1.8	2.2	-0.4 (0.5)
Welfare agency (%)	0.2	0.2	0.0 (0.2)
Mental health agency (%)	0.8	0.8	0.0 (0.3)
Workforce center of unemployment office (%)	0.8	0.9	-0.1 (0.3)
Employer (%)	2.0	2.0	0.0 (0.4)
Other state agency (%)	0.9	1.1	-0.2 (0.4)
Other (%)	1.2	0.8	0.4 (0.4)
Person Who Referred Subject to Training to Learn a New Job or Skill[a]			
Parent/guardian (%)	0.1	0.1	0.0 (0.1)
Spouse/partner (%)	0.1	0.0	0.1 (0.1)
Friend (%)	0.6	0.3	0.3 (0.2)
Job coach (%)	0.1	0.4	-0.3* (0.2)
Employer/supervisor (%)	0.6	0.5	0.0 (0.2)
Other relative (%)	0.2	0.2	0.0 (0.2)
Benefit specialist (%)	0.1	0.1	-0.1 (0.1)
Medical provider (%)	0.1	0.4	-0.3 (0.2)
Other (%)	1.5	1.3	0.1 (0.5)
Not referred by anyone (%)	1.6	1.0	0.6 (0.4)

Source: BOND Stage 1 36-Month Survey.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

[a] The sum of means in this section may exceed the mean reported in the same column in the first row because subjects may have received training from several sources or referral from several sources.

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 (with no multiple-comparisons adjustment).

Exhibit E-7. Estimated Impacts on On-the-Job Training, Job Coaching, or Support Services for Stage 1 Subjects

Outcome	T1 Mean	C1 Mean	Impact Estimate
Received on-the-job training, coaching, or support services (%)	5.7	6.7	-1.0 (1.0)
Source of On-the-Job Training, Coaching, or Support Services[a]			
Vocational rehabilitation agency (%)	2.0	3.0	-0.9 (0.5)
Welfare agency (%)	0.6	0.2	0.3 (0.2)
Mental health agency (%)	1.4	1.4	0.0 (0.4)
Workforce center of unemployment office (%)	0.8	1.5	-0.7 (0.4)
Employer (%)	2.7	3.3	-0.6 (0.5)
Other state agency (%)	1.2	1.0	0.1 (0.4)
Other (%)	0.3	0.5	-0.2 (0.3)
Person Who Referred Subject to Location of On-the-job Training, Coaching, or Support Services[a]			
Parent/guardian (%)	0.2	0.4	-0.1 (0.1)
Spouse/partner (%)	0.0	0.0	0.0 (0.0)
Friend (%)	0.4	0.3	0.1 (0.2)
Job coach (%)	0.2	0.5	-0.3* (0.2)
Employer/supervisor (%)	1.3	1.2	0.1 (0.3)
Other relative (%)	0.1	0.3	-0.2 (0.1)
Benefit specialist (%)	0.1	0.2	-0.1 (0.1)
Medical provider (%)	0.3	0.3	0.0 (0.2)
Other (%)	2.1	2.3	-0.2 (0.6)
Not referred by anyone (%)	0.9	0.7	0.2 (0.3)

Source: BOND Stage 1 36-Month Survey.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

[a] The sum of means in this section may exceed the mean reported in the same column in the first row because subjects may have received training from several sources or referral from several sources.

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 (with no multiple-comparisons adjustment).

Exhibit E-8. Type of Business or Industry and Occupation of the Current Paid Job of Employed Stage 1 Subjects

Outcome	T1 Mean	C1 Mean	Impact Estimate
Type of Business or Industry			
Not currently working for pay or profit (%)	89.4	90.0	-0.6 (1.1)
Natural resources and mining (%)	0.1	0.1	0.0 (0.1)
Goods production (%)	0.8	0.9	-0.1 (0.3)
Trade, transportation, and utilities (%)	2.6	2.4	0.2 (0.5)
Information (%)	0.1	0.3	-0.2** (0.1)
Financial activities (%)	0.2	0.2	0.1 (0.1)
Professional and business services (%)	1.7	1.2	0.5 (0.5)
Education and health services (%)	2.6	3.0	-0.3 (0.6)
Leisure and hospitality (%)	1.7	1.2	0.5 (0.4)
Other services (%)	0.6	0.6	0.0 (0.2)
Public administration (%)	0.2	0.1	0.0 (0.1)
Occupation			
Not currently working for pay or profit (%)	89.4	90.0	-0.5 (1.1)
Management, business, and financial occupations (%)	0.5	0.4	0.1 (0.2)
Computer, engineering, and science occupations (%)	0.1	0.2	-0.1 (0.1)
Community and social services occupations (%)	1.2	0.5	0.7* (0.3)
Health care practitioners and technical occupations (%)	0.1	0.2	-0.1 (0.1)
Service occupations (%)	3.9	3.2	0.8 (0.5)
Sales and related occupations (%)	1.1	1.1	0.1 (0.3)
Office and administrative support occupations (%)	1.5	1.9	-0.3 (0.4)
Farming, fishing, and forestry occupations (%)	0.0	0.0	0.0 (0.0)
Construction and extraction occupations (%)	0.1	0.1	0.0 (0.1)
Installation, maintenance, and repair occupations (%)	0.1	0.1	0.0 (0.1)

Outcome	T1 Mean	C1 Mean	Impact Estimate
Production occupations (%)	0.8	1.3	-0.5 (0.4)
Transportation and material-moving occupations (%)	1.1	1.0	0.1 (0.3)
Military-specific occupations (%)	0.0	0.0	0.0 (0.0)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses and are conditional on employment. Means are regression-adjusted for baseline characteristics. Characteristics of the main job are reported for respondents with more than one job at the time of the survey. Results are pooled for T1 and C1 subjects.

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 (with no multiple-comparisons adjustment).

Exhibit E-9. Special Equipment Used to Help Employed Stage 1 Subjects Work at Current Paid Job

Outcome	T1 Mean	C1 Mean	Impact Estimate
Uses any disability-related special equipment to help work at current job (%)	2.8	2.4	0.5 (0.5)
Type of Equipment Used[a]			
Brace (%)	1.1	0.8	0.4 (0.3)
Cane, crutches, or walker (%)	0.5	0.8	-0.3 (0.2)
Wheelchair (%)	0.3	0.3	0.0 (0.2)
Modified computer hardware (%)	0.2	0.1	0.1 (0.1)
Modified computer software (%)	0.1	0.2	-0.1 (0.1)
Other (%)	1.3	0.6	0.7* (0.4)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses and are conditional on employment. Means are regression-adjusted for baseline characteristics. Characteristics of the main job are reported for respondents with more than one job at the time of the survey. Results are pooled for T1 and C1 subjects.

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

[a] The sum of means in this section may exceed the mean reported in the same column in the first row because subjects may have received training from several sources or referrals from several sources.

*/**/*** 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 (with no multiple-comparisons adjustment).

Exhibit E-10. Estimated Impacts of Employment-Related Outcomes on Stage 1 Subjects at the Time of Survey Interview

Outcome	T1 Mean	C1 Mean	Impact Estimate
Employment Status			
Has worked since random assignment	21.9	22.4	-0.5 (1.9)
Currently working for pay or profit (%)	10.7	10.2	0.5 (1.1)
Currently looking for work in the last four weeks (%)	5.8	6.5	-0.7 (0.9)
Currently not working for pay and not looking for work (%)	83.5	83.3	0.2 (1.4)
Current Job			
Weeks per year working at current main job	4.0	4.0	0.0 (0.4)
Hours per week working at current main job	2.1	2.1	0.1 (0.2)
Has a temporary or seasonal job (%)	2.7	2.3	0.5 (0.7)
Self-employed at current or main job at 36 months after RA (%)	1.2	1.1	0.1 (0.5)
Earnings			
Annual earnings	$1,108	$1,076	$32.58 ($130)
Current weekly earnings above weekly equivalent of BYA (%)	2.8	2.4	0.4 (0.4)
Current weekly earnings above 2 times weekly equivalent of BYA (%)	0.9	0.9	0.0 (0.2)
Current weekly earnings above 3 times weekly equivalent of BYA (%)	0.4	0.5	-0.1 (0.2)
Attitudes Toward Employment			
Personal goals include getting a job (if not working for pay); moving up in a job or learning new job skills (if currently employed) (%)[a]	40.8	39.3	1.5 (2.4)
Personal goals include working and earning enough to stop receiving SSDI (%)[b]	47.1	45.5	1.6 (2.0)
Does volunteer work (%)	9.9	10.9	-1.0 (1.2)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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

[a] Unweighted sample sizes: T1 = 2,916, C1 = 2,819. Sample size is different in this row because the question was asked only of respondents representing themselves and thus did not include responses from proxies.

[b] Unweighted sample sizes: T1 = 2,916, C1 = 2,819. Sample size is different in this row for two reasons: the question was asked only of respondents representing themselves, and the responses include only those respondents who stated that they were receiving Social Security disability benefits at the time of the interview.

*/**/*** 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 (with no multiple-comparisons adjustment).

Exhibit E-11. Estimated Impacts of Employment with Fringe Benefits on Stage 1 Subjects at the Time of Survey Interview

Outcome	T1 Mean	C1 Mean	Impact Estimate
Employment with Fringe Benefits			
Has current job that offers any fringe benefits (%)	5.3	4.5	0.8 (0.6)
Specific Fringe Benefits			
Has current job that offers health care insurance (%)	2.6	2.4	0.2 (0.5)
Has current job that offers dental benefits (%)	2.1	1.8	0.2 (0.4)
Has current job that offers sick days with pay (%)	2.0	2.0	0.0 (0.5)
Has current job that offers long-term disability benefits (%)	1.0	1.0	0.0 (0.2)
Has current job that offers short-term disability benefits (%)	1.3	1.3	0.0 (0.3)
Has current job that offers flexible health or dependent care spending accounts (%)	0.8	0.7	0.0 (0.3)
Has current job that offers paid vacation (%)	3.0	2.8	0.2 (0.4)
Has current job that offers free or low-cost child care (%)	0.1	0.1	0.0 (0.1)
Has current job that offers transportation assistance (%)	1.2	1.0	0.2 (0.3)
Has current job that offers pension or retirement benefits (%)	2.2	1.6	0.5 (0.4)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

Exhibit E-12. Estimated Impacts of Employment-Related Expenses on Stage 1 Subjects at the Time of Survey Interview

Outcome	T1 Mean	C1 Mean	Impact Estimate
Total weekly work-related expenses	$3.46	$3.07	$0.39 ($0.60)
Specific Work-Related Expenses			
Weekly work-related commuting expenses	$2.64	$2.50	$0.14 ($0.51)
Weekly work-related expenses not including commute, such as uniforms, licenses, permits, union dues, special tools	$0.20	$0.29	-$0.09 ($0.07)
Weekly work-related child care expenses	$0.55	$0.18	$0.38* ($0.20)
Weekly work-related expenses for special equipment to accommodate disability, such as a brace, cane, or modified computer software, not reimbursed or covered by insurance	$0.06	$0.07	-$0.01 ($0.03)
Weekly work-related expenses for personal assistance services related to disability, such as a sign language interpreter or personal care attendant, not reimbursed or covered by insurance	$0.02	$0.06	-$0.04 ($0.03)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

Exhibit E-13. Estimated Impacts of Employer Accommodations on Stage 1 Subjects at the Time of Survey Interview

Outcome	T1 Mean	C1 Mean	Impact Estimate
Received any accommodation from employer (%)	6.8	6.6	0.2 (0.9)
Specific Employer Accommodations			
Employer provided special equipment or assistive technology (%)	0.9	0.6	0.3 (0.2)
Employer kept job available during disability-related absences (%)	4.4	4.9	-0.5 (0.7)
Employer arranged for coworkers to help subject, when needed (%)	5.8	5.5	0.3 (0.8)
Employer provided modified computer hardware (%)	0.4	0.2	0.2 (0.1)
Employer provided modified computer software (%)	0.3	0.1	0.2* (0.1)
Employer made other accommodations not listed above (%)	0.2	0.3	-0.1 (0.1)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

Exhibit E-14. Estimated Impacts of Household Income and Material Hardship Outcomes on Stage 1 Subjects at the Time of Survey Interview

Outcome	T1 Mean	C1 Mean	Impact Estimate
Household Income			
Household income in 2013[a]	$25,004	$25,028	-$23.88 ($1,097)
Living Below Poverty Line			
Household income below federal poverty line in 2013[a] (%)	44.3	46.2	-1.9 (2.0)
Could Not Meet Household Expenses			
Any time in past 12 months when did not meet all essential expenses (%)	49.3	51.2	-1.9 (1.8)
Did not pay full amount of rent or mortgage (any time in past 12 months) (%)	24.0	24.4	-0.4 (1.6)
Evicted for not paying rent or mortgage (any time in past 12 months) (%)[b]	1.9	1.8	0.2 (0.5)
Could not pay full amount of utility bills (any time in past 12 months) (%)	36.7	40.0	-3.3* (1.8)
Utility turned off service because of nonpayment (any time in past 12 months) (%)[c]	8.4	8.5	-0.1 (1.1)
Telephone or cell company disconnected because of nonpayment (any time in past 12 months) (%)	22.2	21.3	1.0 (1.5)
Food Security			
Low food security (%)[d]	22.5	21.0	1.6 (2.6)
Very low food security (%)[d]	29.7	32.6	-2.9 (2.1)
Food security scale[e]	6.2	6.4	-0.2 (0.1)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

[a] The survey question asked respondents to indicate the total combined income of all members of the household during the last calendar year. Survey data collection began in May 2014 and ended in February 2015. For the large majority of sample respondents, the survey question asked about 2013 income.

[b] This question was asked only of respondents who reported that they faced difficulty in paying for their household expenses in past 12 months, which was nn.n% of the weighted responses. The means and impact estimates in this row reflect responses from unconditional responses. Respondents who did not face any difficulty in paying for their household expenses are treated as zeros.

[c] This question was asked only of respondents who reported that they could not pay the full amount of gas, oil, or electricity bills in past 12 months, which was nn.n% of the weighted responses. The means and impact estimates in this row reflect responses from unconditional responses. Respondents who did not face any difficulty in paying the full amount of gas, oil, or electricity bills are treated as zeros.

[d] Low food security and very low food security are mutually exclusive categories. Low food security includes beneficiaries who respond as indicated to two to four of the following, as assessed over the past 12 months: (1) often or sometimes ran out of food and didn't have money to get more, (2) often or sometimes couldn't afford to eat balanced meals, (3) cut or skipped meals for financial reasons, (4) cut or skipped meals for financial reasons in almost every month or some months but not every month, (5) ate less than desired for financial reasons, (6) was hungry but didn't eat for financial reasons. Very low food security includes beneficiaries who respond as indicated for five or six of the questions.

[e] The food security scale comes from the U.S. Department of' Agriculture U.S. Household Food Security Survey Module: Six-Item Short Form Survey. The minimum value is 0, the maximum value is 8.48. The higher the score, the less food secure.

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 (with no multiple-comparisons adjustment).

Exhibit E-15. Estimated Impacts of Self-Reported Benefit Receipt on Stage 1 Subjects at the Time of Survey Interview

Outcome	T1 Mean	C1 Mean	Impact Estimate
Receipt of any supports (excluding SSDI) in the last month (%)	54.2	52.9	1.3 (1.8)
Specific Benefits Received			
Receipt of SSDI in the last month (%)	97.8	97.4	0.4 (0.4)
Receipt of SNAP (Food Stamps) in the last month (%)	33.3	34.1	-0.8 (1.6)
Receipt of public assistance/welfare (TANF) benefits in the last month (%)	2.9	3.7	-0.8 (0.7)
Receipt of veterans' benefits in the last month (%)	3.0	3.6	-0.7 (0.7)
Receipt of workers' compensation benefits in the last month (%)	1.9	2.6	-0.7 (0.7)
Receipt of private disability insurance benefits in the last month (%)	3.1	2.4	0.7 (0.6)
Receipt of disability insurance for a disabled adult child in the last month (%)	1.0	1.0	0.0 (0.3)
Receipt of unemployment insurance benefits in the last month (%)	0.2	0.2	0.0 (0.2)
Receipt of private pensions or government employee pensions in the last month (%)	6.0	5.8	0.3 (0.8)
Receipt of other government assistance (for example, energy assistance or child care assistance) (%)	7.0	6.8	0.2 (1.0)
Receipt of other assistance on a regular basis (%)	5.7	5.2	0.4 (0.9)
Receipt of other assistance not on a regular basis (%)	2.7	2.9	-0.2 (0.7)
Receipt of government housing assistance in the last month (%)	11.4	8.8	2.6 (1.5)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

Exhibit E-16. Estimated Impacts of Self-Reported Benefit Levels on Stage 1 Subjects at the Time of Survey Interview

Outcome	T1 Mean	C1 Mean	Impact Estimate
Total benefit amount in the last month (excluding SSDI)	$215	$223	-$8.29 ($22.03)
Specific Benefit Levels			
SNAP (Food Stamps) in the last month	$48.06	$48.32	-$0.26 ($3.50)
Public assistance/welfare (TANF) benefits in the last month	$3.82	$4.87	-$1.05 ($1.32)
Veterans' benefits in the last month	$37.30	$40.35	-$3.06 ($14.04)
Workers' compensation benefits in the last month	$17.95	$28.17	-$10.21 ($11.09)
Private disability insurance benefits in the last month	$15.76	$24.09	-$8.33 ($7.95)
Disability insurance for a disabled adult child in the last month	$3.49	$4.27	-$0.78 ($1.71)
Unemployment insurance (UI) benefits in the last month	$0.54	$0.59	-$0.06 ($0.68)
Private pensions or government employee pensions in the last month	$44.23	$32.69	$11.54 ($7.57)
Other government assistance (for example, energy assistance or child care assistance)	$10.95	$12.31	-$1.36 ($3.05)
Other assistance on a regular basis in the last month	$28.33	$22.45	$5.87 ($7.87)
Other assistance not on a regular basis in the last month	$6.27	$7.32	-$1.05 ($3.03)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

Exhibit E-17. Estimated Impacts of Living Situation on Stage 1 Subjects at the Time of Survey Interview

Outcome	T1 Mean	C1 Mean	Impact Estimate
Number of people in the household	1.46	1.44	0.02 (0.06)
Single-family home (%)	58.7	57.3	1.4 (1.8)
Mobile home (%)	10.1	9.5	0.6 (1.1)
Regular apartment (%)	24.6	25.2	-0.6 (1.6)
Supervised apartment (%)	1.4	1.2	0.3 (0.4)
Group home (%)	1.9	2.6	-0.7 (0.5)
Halfway house (%)	0.2	0.1	0.1 (0.2)
Personal care or board and care home (%)	0.3	0.2	0.1 (0.2)
Assisted living facility (%)	1.0	1.1	-0.1 (0.4)
Nursing or convalescent home (%)	0.5	0.8	-0.3 (0.4)
Shelter (%)	0.1	0.3	-0.2 (0.2)
Other supervised group residence or facility (%)	0.3	0.4	-0.1 (0.2)
Other (%)	0.9	1.2	-0.3 (0.4)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

Exhibit E-18. Estimated Impacts of Health Outcomes on Stage 1 Subjects at the Time of Survey Interview: T1 Versus C1

Outcome	T1 Mean	C1 Mean	Impact Estimate
Health Status (global)			
Health is excellent (%)	3.3	2.7	0.6 (0.6)
Health is very good (%)	5.8	6.7	-0.9 (0.8)
Health is good (%)	20.6	20.9	-0.3 (1.5)
Health is fair (%)	38.3	37.2	1.2 (1.8)
Health is poor (%)	32.0	32.5	-0.5 (1.7)
Physical Health and Functioning			
Composite physical health score (SF physical component summary)[a, b]	34.2	33.9	0.3 (0.5)
Composite mental health score (SF mental component summary) [a, b]	41.9	41.4	0.5 (0.5)
Stayed overnight in the hospital in the last 12 months (%)	30.2	31.7	-1.5 (2.3)
Number of nights in the hospital in the last 12 months	4.1	4.2	-0.1 (0.8)
Number of days in the last 12 months when illness or injury kept subject in bed for more than half the day	49.9	55.2	-5.3 (3.5)
Needs help with personal care such as bathing, dressing (%)	26.1	27.6	-1.5 (1.6)
Needs the help of another person in order to get around inside own home (%)	15.3	17.9	-2.6 (1.4)
Needs the help of another person in order to get around outside own home (%)	30.9	33.6	-2.7 (2.0)
Emotional Health and Functioning			
Has a lot of trouble concentrating long enough to finish everyday tasks (%)[b]	46.9	48.8	-1.9 (1.8)
Has a lot of trouble coping with day-to-day stress (%)[b]	47.7	49.8	-2.1 (1.8)
Overweight and Obesity[c]			
Underweight (%)	2.9	2.4	0.4 (0.6)
Overweight (%)	27.7	27.5	0.2 (2.1)
Obese (%)	45.5	46.4	-0.9 (2.0)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

[a] Optum SF Health Survey Scale. The scale ranges from 0 to 100. The higher the score, the healthier. Mean scale score in the whole adult U.S. population is 50.

[b] Unweighted sample sizes: T1 = 2,916, C1 = 2,819. Sample size is different in this row because this question was asked only of respondents representing themselves and thus did not include responses from proxies.

[c] Underweight is defined as a Body Mass Index (BMI) under 18.5. Overweight is defined as a BMI between 25.0 and 29.9. Obese is defined as a BMI of 30.0 or greater.

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 (with no multiple-comparisons adjustment).

Exhibit E-19. Estimated Impacts of Health Insurance on Stage 1 Subjects at the Time of Survey Interview: T1 Versus C1

Outcome	T1 Mean	C1 Mean	Impact Estimate
Has health insurance coverage (%)	96.8	97.1	-0.3 (0.8)
Insured by Medicaid (%)	37.7	38.8	-1.2 (1.8)
Insured by Medicare (%)	78.7	78.0	0.8 (1.5)
Insured by CHAMPUS/CHAMP-VA, Veterans Affairs, or other military health insurance program (%)	3.9	3.4	0.5 (0.7)
Insured by Indian Health Service (%)	0.2	0.0	0.2 (0.1)
Insured by Medi-GAP (%)	0.3	0.0	0.3 (0.2)
Insured by state program (%)	3.3	3.0	0.3 (0.7)
Has private insurance through own employer (%)	3.6	4.2	-0.6 (0.7)
Has private insurance through spouse/partner/family employer (%)	9.8	7.3	2.4 (1.5)
Has private insurance paid by self or family (%)	6.4	5.6	0.7 (1.0)
Has private disability insurance paid by self or family (%)	0.5	0.3	0.1 (0.3)
Has other insurance plan (%)	8.3	9.8	-1.5 (1.7)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

Exhibit E-20. Estimated Impacts of Marital Status on Stage 1 Subjects at Time of Interview: T1 Versus T2

Outcome	T1 Mean	C1 Mean	Impact Estimate
Married (%)	30.2	30.3	-0.1 (1.6)
Widowed (%)	5.5	5.1	0.4 (1.0)
Divorced (%)	23.0	23.9	-0.8 (1.7)
Separated (%)	6.0	5.5	0.5 (0.9)
Never married (%)	35.3	35.2	0.1 (1.4)
Currently living with a spouse or someone like a spouse (%)	34.6	35.7	-1.1 (1.7)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

Exhibit E-21. Estimated Impacts of Time Use on Stage 1 Participants at the Time of Survey Interview

Outcome	Treatment Mean	Control Mean	Impact Estimate
Hours per week spent on measured activities	23.2	22.4	0.8 (0.9)
Specific Time Use			
Hours per week working in a job for pay	2.5	2.5	-0.1 (0.2)
Hours per week commuting to and from work	0.7	0.6	0.1 (0.1)
Hours per week doing unpaid work at family business	0.3	0.3	-0.1 (0.1)
Hours per week volunteering for an organization	0.7	0.9	-0.2 (0.2)
Hours per week in school, working toward a degree program, or in a training program	0.7	0.5	0.2 (0.2)
Hours per week homemaking, including caring for others, food preparation, yard work, and house repairs	12.1	11.5	0.6 (0.7)
Hours per week devoted to personal health care and self-grooming	8.0	7.7	0.3 (0.4)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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 (with no multiple-comparisons adjustment).

Exhibit E-22. Estimated Impacts on Employment-Related Outcomes of Stage 1 Subjects at the Time of Survey Interview for Subgroups Defined by Employment in 2010

Outcome	Employed			Not Employed			Estimated Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Employment Status							
Has worked since random assignment	61.7	61.8	-0.1 (3.5)	12.0	12.6	-0.6 (1.6)	0.5 (3.9)
Currently working for pay or profit (%)	34.5	34.3	0.2 (2.7)	5.7	5.1	0.6 (1.0)	-0.4 (2.9)
Currently looking for work in the last four weeks (%)	11.2	13.0	-1.7 (2.4)	4.7	5.1	-0.5 (0.9)	-1.3 (2.5)
Currently not working for pay and not looking for work (%)	54.3	52.7	1.6 (3.1)	89.7	89.8	-0.1 (1.3)	1.7 (3.4)
Earnings							
Annual earnings	$3,991	$3,901	$90.19 ($419)	$513	$492	$20.71 ($113)	$69.48 ($434)
Current weekly earnings above weekly equivalent of BYA (%)	9.9	7.6	2.2 (1.3)	1.3	1.3	0.0 (0.3)	2.3 (1.4)
Current weekly earnings above 2 times weekly equivalent of BYA (%)	3.1	3.8	-0.7 (0.9)	0.4	0.3	0.1 (0.2)	-0.8 (0.9)
Current weekly earnings above 3 times weekly equivalent of BYA (%)	1.4	2.1	-0.6 (0.5)	0.2	0.2	0.0 (0.2)	-0.6 (0.6)

Source: BOND Stage 1 36-Month Survey and baseline SSA administrative data.

Notes: Weights reflecting sample selection and survey nonresponse ensure that the BOND subjects who met analysis criteria are representative of SSDI recipients in the nation. Standard errors appear in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

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

[a] Unweighted sample sizes: T1 = 2,916, C1 = 2,819. Sample size is different in this row because the question was asked only asked of respondents representing themselves and thus did not include responses from proxies.

[b] Unweighted sample sizes: T1 = 2,916, C1 = 2,819. Sample sizes are different in this row for two reasons: the question was asked only of respondents representing themselves and only of respondents who stated that they were receiving Social Security disability benefits at the time of interview.

*/**/*** 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 (with no multiple-comparisons adjustment).

Exhibit E-23. Summary of Statistically Significant Impact Estimates from the Stage 1 36-Month Survey: Employment-Related Domains

Outcome Domain	Statistically Significant	Not Statistically Significant
Type of business or industry	Information	Natural resources and mining; goods production; trade, transportation, and utilities; financial activities; professional and business services; education and health services; leisure and hospitality; other services; public administration
Occupation	Community and social services occupations	Management, business, and financial occupations; computer, engineering, and science occupations; health care practitioners and technical occupations; sales and related occupations; farming, fishing, and forestry occupations; construction and extraction occupations; installation, maintenance, and repair occupations; transportation and material-moving occupations; military-specific occupations; service occupations; office and administrative support occupations; production occupations
Special equipment for employment		Uses any disability-related special equipment to help work at current job; brace; cane, crutches, or walker ; wheelchair; modified computer hardware; modified computer software; other
Employment status		Has worked since random assignment; current work status
Current job		Weeks per year working at current job Hours per week working at current job; has a temporary or seasonal job Self-employed at current or main job
Earnings		Annual earnings; current weekly earnings in excess of 1 times BYA weekly equivalent; current weekly earnings in excess of 2 times BYA weekly equivalent; current weekly earnings in excess of 3 times BYA weekly equivalent
Attitudes toward employment		Personal goals include getting a job; personal goals include earning enough to stop receiving SSDI; does volunteer work
Employment with fringe benefits		Current job offers any fringe benefits; has current job that offers health care insurance; has current job that offers dental benefits; has current job that offers sick days with pay; has current job that offers long-term disability benefits; has current job that offers short-term disability benefits; has current job that offers flexible health or dependent care spending accounts; has current job that offers paid vacation; has current job that offers free or low-cost child care; has current job that offers transportation assistance; has current job that offers pension or retirement benefits
Work-related expenses	Work-related child care expenses	Total work-related expenses; weekly work-related commuting expenses; weekly work-related expenses not including commute, such as uniforms, licenses, permits, union dues, special tools; weekly work-related expenses for special equipment to accommodate disability; weekly work-related expenses for personal assistance services related to disability not reimbursed or covered by insurance
Employer accommodations		Received any accommodation from employer; employer provided special equipment or assistive technology; employer kept job available during disability-related absences; employer arranged for coworkers to help subject, when needed; employer provided modified computer hardware; employer provided modified computer software; employer made other accommodations not listed above

Notes: The first column describes the outcome domain, the second column identifies which outcomes in the domain had statistically significant impacts, and the third column identifies which outcomes in the domain had statistically insignificant impacts.

Exhibit E-24. Summary of Statistically Significant Impact Estimates from the Stage 1 36-Month Survey: Household- and Income-Related Domains

Outcome Domain	Statistically Significant	Not Statistically Significant
Household income		2012 household income; household income below poverty line
Could not meet household expenses	Could not pay full amount of utility bills	Did not meet all essential expenses; did not pay full amount of rent or mortgage; evicted for not paying rent or mortgage; asked utility to turn off service; asked telephone company to disconnect because of nonpayment
Food security		Low food security; very low food security; food security scale
Benefit receipt		Receipt of any supports (excluding SSDI) in the last month; receipt of SSDI in the last month; receipt of SNAP in the last month; receipt of public assistance/welfare (TANF) benefits in the last month; receipt of veterans' benefits in the last month; receipt of workers' compensation benefits in the last month; receipt of private disability insurance benefits in the last month; receipt of disability insurance for a disabled adult child in the last month; receipt of UI benefits in the last month; receipt of private pensions or government employee pensions in the last month; receipt of other government assistance (for example, energy assistance or child care assistance); receipt of other assistance on a regular basis; receipt of other assistance not on a regular basis; receipt of government housing assistance in the last month
Benefit levels		Total benefit amount in the last month (excluding SSDI); SNAP in the last month; public assistance/welfare (TANF) benefits in the last month; Veterans' benefits in the last month; workers' compensation benefits in the last month; private disability insurance benefits in the last month; disability insurance for a disabled adult child in the last month; UI benefits in the last month; private pensions or government employee pensions in the last month; other government assistance (for example, energy assistance or child care assistance); other assistance on a regular basis in the last month; other assistance not on a regular basis in the last month
Living situation		Number of people in the household; single-family home; mobile home; regular apartment; supervised apartment; group home; halfway house; personal care or board and care home; assisted living facility; nursing or convalescent home; shelter; other supervised group residence or facility; other

Notes: The first column describes the outcome domain, the second column identifies which outcomes in the domain had statistically significant impacts, and the third column identifies which outcomes in the domain had statistically insignificant impacts.

Exhibit E-25. Summary of Statistically Significant Impact Estimates from the Stage 1 36-Month Survey: Health-Related Domains

Outcome Domain	Statistically Significant	Not Statistically Significant
Health status		Health is excellent; health is very good; health is good; health is fair; health is poor
Physical health and functioning		Composite physical health score; composite mental health score; stayed overnight in the hospital in the last 12 months; number of nights in the hospital in the last 12 months; number of days in the last 12 months when illness or injury kept subject in bed for more than half the day; needs help with personal care such as bathing, dressing; needs the help of another person in order to get around inside own home; needs the help of another person in order to get around outside own home
Emotional health and functioning		Trouble concentrating long enough to finish everyday tasks; trouble coping with day-to-day stress
Weight status		Underweight; overweight; obese
Health insurance		Has health insurance coverage; insured by Medicaid; insured by Medicare; insured by CHAMPUS/CHAMP-VA, the Department of Veterans Affairs, or other military health insurance program; insured by Indian Health Service; insured by Medi-GAP; insured by state program; has private insurance through own employer; has private insurance through spouse/partner/family employer; has private insurance paid by self or family; has private disability insurance paid by self or family; has other insurance plan
Marital status		Married; widowed; divorced; separated; never married; currently living with a spouse or someone like a spouse
Time use		Hours per week spent on measured activities; hours per week working in a job for pay; hours per week commuting to and from work; hours per week doing unpaid work at family business; hours per week volunteering for an organization; hours per week in school, working toward a degree, or in a training program; hours per week homemaking, including caring for others, food preparation, yard work, and house repairs; hours per week devoted to personal health care and self-grooming

Notes: The first column describes the outcome domain, the second column identifies which outcomes in the domain had statistically significant impacts, and the third column identifies which outcomes in the domain had statistically insignificant impacts.