

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these estimates have been updated over the course of the demonstration, making the published estimates in this report out of date.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

First- and Second-Year Snapshot of Earnings and Benefit Impacts for Stage 2

Deliverable 24c.5

August 14, 2014

Submitted To:

Social Security Administration

Attn: Ms. Joyanne Cobb

Office of Research, Demonstrations, and
Employment Support

500 E Street, SW

9th Floor, RM # 905

Washington, DC 20254

Contract No. SS00-10-60011



Prepared by

Daniel Gubits

Winston Lin

Stephen Bell

David Judkins

Abt Associates

55 Wheeler Street

Cambridge, MA 02138-1168

Report Context

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 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. 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 more about impacts for those most likely to use the offset (recruited and informed volunteers) and to determine the extent to which Enhanced Work Incentives Counseling (EWIC) affects impacts.

This report, the first of two Stage 2 *Snapshot Reports*, documents Stage 2 impacts on earnings and benefit outcomes during the first and second calendar years of implementation (2011 and 2012). The Stage 2 sample of SSDI-only volunteers includes the offset-only group (T21), the offset-EWIC group (T22), and the Stage 2 control group (C2). The authors have conducted pairwise comparisons of outcomes for beneficiaries in these groups to provide estimates of the impact of the offset compared to current law (T21 vs. C2); the impact of the offset plus EWIC, again compared to current law (T22 vs. C2); and the marginal impact of EWIC once the offset is available to both groups (T22 vs. T21).

Future reports--Interim Process, Participation, and Impact reports in 2015 and 2017, a second Snapshot report in 2016 and the final report in 2017--will track Stage 2 impacts through 2015. A parallel series of reports is being produced for Stage 1, the first two of which were released in 2013 and 2014.

Summary of Key Findings

The findings in this report concern the SSDI beneficiaries thought most likely to respond to the offset work incentives: volunteers who wished to have the offset rules applied to them and who did not initially receive SSI. The report focuses on the second calendar year of Stage 2 (2012), which is the

first year in which the offset and its work incentives counseling components could have affected outcomes for all Stage 2 subjects. The impact estimates show that:

- There is some confirmatory evidence that the offset rules combined with standard work incentives counseling (WIC) *increased mean earnings and SSDI benefits paid* compared to current law earnings rules and counseling services. There is also suggestive evidence (from non-confirmatory tests) of increases in the proportions of beneficiaries with any earnings and with earnings above the BOND Yearly Amount (BYA).
- There is some confirmatory evidence that the offset rules combined with enhanced work incentives counseling (EWIC) *increased mean earnings* when compared to current law. There is also suggestive evidence of increases in the proportions of beneficiaries with earnings above BYA and above 3 x BYA, as well as an increase in the number of months with SSDI payments.
- When combined with the offset, EWIC *did not have any detectable effects on 2012 earnings and benefit outcomes* relative to WIC.

The BOND Evaluation Team

Abt Associates, in partnership with 25 other organizations, is implementing and evaluating the BOND under contract to the U.S. Social Security Administration. To ensure the objectivity of the evaluation, separate teams conduct the implementation and evaluation 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. These individuals have no role in implementing or overseeing the BOND intervention they are studying, nor do any members of their evaluation team. 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.

Table of Contents

Acronyms Used in This Report.....	ii
Terminology.....	iii
1. Introduction.....	1
1.1. Current SSDI Rules and the BOND Innovation.....	1
1.2. BOND Stage 2 Implementation and Random Assignment	2
1.3. Purpose	4
1.4. Organization of the Report	5
2. Methodology and Context	6
2.1. Outcome Definitions and Theoretical Impacts.....	7
2.2. Administrative Features of the Offset That Influence Impacts	8
2.3. Impact Estimation Methodology	10
2.4. Final Analysis Sample Sizes	11
3. Impact Findings.....	13
3.1. Impacts for Full Stage 2 Sample	14
3.1.1. Confirmatory Impacts.....	14
3.1.2. Exploratory Impacts	14
3.2. Subgroups.....	16
3.2.1. Subgroups Defined by Duration of SSDI Participation	16
3.2.2. Subgroups Defined by Year of Demonstration Enrollment	16
4. Discussion.....	23
References.....	24
Appendix A. Stage 2 Impact Estimates on 2011 Earnings and Benefit Outcomes	25
Appendix B. Detailed Summary of Methodological Approach.....	27
B.1 Changes from the <i>Evaluation Analysis Plan</i> and from Stage 1 Impact Estimation	27
B.2 Estimation Procedure	28
B.3 Multiple Comparisons.....	30
B.4 Covariates.....	32
B.5 Analysis Weights and Sample Exclusions	36
B.5.1. Construction of Analysis Weights.....	36
B.5.2. Sample Exclusions	38

Acronyms Used in This Report

AEE	Annual Earnings Estimate
BODS	BOND Operations Data System
BOND	Benefit Offset National Demonstration
BYA	BOND Yearly Amount (equal to $12 \times$ the monthly SGA level)
DAC	Disabled Adult Child
DWB	Disabled Widow/Widowers Benefits
EWIC	Enhanced Work Incentives Counseling
GP	Grace Period
IRS	Internal Revenue Service
IRWE	Impairment Related Work Expenses
MEF	Master Earnings File
PHUS	Payment History Update system
SGA	Substantial Gainful Activity
SSA	Social Security Administration
SSDI	Social Security Disability Insurance
SSI	Supplemental Security Income
SSR	Supplemental Security Record
TWP	Trial Work Period
WIC	Work Incentives Counseling

Terminology

1. **Prospective BOND subjects:** beneficiaries in the pool eligible for potential assignment at Stage 1.
2. **Stage 2 solicitation pool:** SSDI-only beneficiaries to be recruited for Stage 2.
3. **Stage 2 volunteers:** those subjects who volunteer for Stage 2.
4. **BOND subjects:** beneficiaries assigned to any of the five BOND treatment or control groups, at either stage (see **Exhibit 2-3**). Terms for subjects in specific groups are as follows:
 - a. **Treatment subjects:** All subjects offered the use of the benefit offset, including:
 - i. **T1 subjects** or **Stage 1 treatment subjects:** Those offered the offset at Stage 1.
 - ii. **Stage 2 treatment subjects:** Those offered the offset at Stage 2, including:
 - (1) **T21 subjects** or **Stage 2 offset-only subjects:** Stage 2 volunteers offered the offset, but not offered enhanced work incentives counseling.
 - (2) **T22 subjects** or **Stage 2 offset-EWIC subjects:** Stage 2 volunteers offered both the offset and enhanced work incentives counseling.
 - b. **Control subjects:** Those whose benefits will continue to be determined by current law.
 - i. **C1 subjects** or **Stage 1 control subjects:** Those assigned to the Stage 1 control group.
 - ii. **C2 subjects** or **Stage 2 control subjects:** Stage 2 volunteers assigned to the Stage 2 control group.
5. **BOND users:** those treatment subjects who take up a BOND treatment. These include:
 - a. **Offset-only users** – all treatment subjects who have their benefits reduced by the offset but *do not use EWIC*, either because EWIC is not offered or because they choose not to avail themselves of it.
 - b. **EWIC-only users** – all treatment subjects who use EWIC services *but do not have their benefits reduced by the offset*, because their earnings never rise high enough to use it. They can only be subjects in the T22 group.
 - c. **Offset - EWIC users** – All treatment subjects who use EWIC services *and have their benefits reduced by the offset*. They can only be subjects in the T22 group.
 - d. **Offset users** – the combination of offset-only and offset-EWIC users.
 - e. **EWIC users** – the combination of EWIC-only and offset-EWIC users.

1. Introduction

The Benefit Offset National Demonstration (BOND) is a random assignment demonstration that tests a variant of Social Security Disability Insurance (SSDI) program rules governing work and other supports. This Snapshot Report concerns Stage 2 of BOND, which was designed to learn about the impacts of the benefit offset for those most likely to use it, and to determine the marginal effects of the delivery of more intensive counseling services than those offered under current law. This is the first of two *Stage 2 Snapshot Reports* about these innovations' impacts on earnings and benefits paid. This introductory chapter describes the benefit offset and Stage 2 of the demonstration, explains the purpose of this report, and ends with an outline of the remainder of the report.

1.1. Current SSDI Rules and the BOND Innovation

Under current program rules, SSDI beneficiaries lose all SSDI benefits after a sustained period of substantial earnings and risk potential loss of other (non-SSDI) benefits.¹ Specifically, benefits are lost if an SSDI beneficiary's 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 2012, the SGA amount was \$1,010 per month for non-blind beneficiaries and \$1,690 per month for blind beneficiaries. The complete loss of benefits for earnings in excess of the SGA amount is sometimes called the "cash cliff." The cash cliff likely discourages some beneficiaries from working at all and encourages those who do work to keep their earnings below the SGA level.

BOND replaces the cash cliff with a "ramp" (i.e., the benefit offset) with the policy objective of encouraging beneficiaries to increase their earnings and reduce their reliance on benefits.² The benefit offset is expected to increase the earnings of those who might otherwise not work at all and those who already work but might not attempt to earn more than the SGA amount. If such individuals engage in SGA under the benefit offset, their benefits will eventually be reduced. Ultimately, however, the direction of net impacts on mean earnings and benefits for 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. Those who would engage in SGA even without the incentive of the offset would lose their benefits entirely under current law, whereas under the benefit offset many, perhaps most, might be eligible for a reduced SSDI benefit. While still on the ramp—i.e., while earning above the SGA amount but less than the zero-benefit amount at the end of the ramp—beneficiaries can increase the size of their benefits by working at less than their full earnings potential.

Differences between the administration of the benefit offset and the administration of current law benefits are also expected to contribute to impacts, especially the impact on benefits. One of these differences is that BOND uses an annual accounting period, rather than the monthly period used under current law. SSA continues to pay benefits monthly, but the monthly payment amount is initially based on an estimate of

¹ Other benefits include Medicare for those on the rolls for at least 24 months, which are extended for a lengthy period following suspension of SSDI benefits, but not indefinitely. Some beneficiaries also receive Supplemental Security Income, Medicaid, or a variety of other public or private benefits that are contingent on earnings in some fashion.

² See Exhibit 1-1 of the *Stage 2 Early Assessment Report* (Gubits et al., 2013) for a detailed comparison of current SSDI program rules with BOND rules related to work.

calendar-year earnings.³ SSA's end-of-year reconciliation (to IRS earnings records) process leads to adjustments if actual earnings deviate more than \$200 from the beneficiary's estimate. The benefit offset reduces benefits by \$1 for every \$2 in countable annual earnings in excess of the BOND Yearly Amount (BYA). BYA is equal to 12 times the monthly SGA amount. In 2012, BYA was \$12,120 for non-blind and \$20,280 for blind subjects. The change to an annual accounting period was designed to reduce the cost of administering the offset. It can also be very helpful to 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.

1.2. BOND Stage 2 Implementation and Random Assignment

To support rigorous estimation of the impacts of offering the benefit offset to the SSDI beneficiary population, the design of BOND was developed in two segments referred to as "Stage 1" and "Stage 2" of the demonstration. Stage 1 was designed to examine how a national benefit offset and accompanying administrative changes would affect earnings and program outcomes for the entire SSDI population nationally. Stage 2 was designed to learn more about the impacts of the benefit offset for those most likely to use it (recruited and informed volunteers) and to determine the extent to which significant enhancements to the basic BOND-focused work incentives counseling affect offset utilization and impacts.

For the demonstration, SSA randomly selected ten large study sites to statistically represent the nation.⁴ Eligible beneficiaries in those sites were first assigned at random to a Stage 1 offset-only treatment group, a Stage 1 control group, or a pool to be solicited as volunteers for Stage 2.⁵ Those eligible for the two Stage 1 groups included both SSDI-only and concurrent (SSDI and SSI) beneficiaries, while eligibility for Stage 2 is limited to SSDI-only beneficiaries.⁶ Of those beneficiaries who were solicited to participate in the Stage 2 study, about 5 percent volunteered for the study.⁷ Those who volunteered were then randomly assigned to one of the three groups:

- **T21 subjects** (Stage 2 offset-only subjects): a group that receives the \$1 for \$2 benefit offset with Work Incentives Counseling (WIC) only;

³ Beneficiaries may submit revised earnings estimates during the year if their income deviates from their initial estimates.

⁴ As explained in Chapter 2, findings on BOND impacts from Stage 1 of the demonstrations (see Stapleton et al., 2012, 2014) apply to the nation as a whole, while those from Stage 2 in the current report apply to just beneficiaries in the ten study sites. The detailed statistical reason for this property of the Stage 2 findings is provided in Appendix B.

⁵ The Stage 1 impact analysis compares outcomes of the Stage 1 treatment group with outcomes of the Stage 1 control group.

⁶ Concurrent beneficiaries were excluded from Stage 2 because the interaction between SSDI and SSI rules substantially diminishes the value of the SSDI offset to concurrent beneficiaries, leading to an expectation that relatively few concurrent beneficiaries would use the SSDI benefit offset.

⁷ The Stage 2 outreach is described in detail in the *Stage 2 Early Assessment Report* (Gubits et al., 2013).

- **T22 subjects** (Stage 2 offset-EWIC subjects): a group that receives the \$1 for \$2 benefit offset and Enhanced Work Incentives Counseling (EWIC); or
- **C2 subjects** (Stage 2 control subjects): a control group that is not offered the offset or EWIC and is subject to current law.

Random assignment for Stage 2 occurred between March 1, 2011 and September 28, 2012, with 40 percent of volunteers enrolling in the study in 2011 and 60 percent of volunteers enrolling in 2012. In total, 12,954 beneficiaries were randomly assigned to the three groups. The random assignment ratio for the three assignment groups was 8:5:8; ultimately, 4,935 volunteers were assigned to the T21 group, 3,089 volunteers were assigned to the T22 group, and 4,930 volunteers were assigned to the C2 group.

It was expected that beneficiaries who had received SSDI for a short duration (defined as three years or less) would be more likely to work—and so be more responsive to the work incentives in BOND—than beneficiaries who had received SSDI for a longer time. Therefore, SSA especially sought information in Stage 2 for beneficiaries who had received SSDI for a short duration. To do this, it set a goal of having at least 50 percent of volunteers be short-duration recipients. Because only 32 percent of SSDI-only beneficiaries overall fall into this subpopulation, this goal was accomplished by oversampling short-duration beneficiaries into the Stage 2 outreach waves. Short-duration beneficiaries make up 64 percent of all Stage 2 subjects.⁸

The impact analysis for Stage 2 includes three pairwise comparisons which serve to address three research questions:

Research Question	Addressed by Comparison of
A. What is the <i>impact of the benefit offset on SSDI-only beneficiaries who volunteer for BOND</i> , compared to current law?	T21 to C2
B. What is the <i>impact of the benefit offset plus enhanced work incentives counseling</i> on SSDI-only beneficiaries who volunteer for BOND, compared to current law?	T22 to C2
C. What is the <i>incremental effect of enhanced work incentives counseling when added to the benefit offset</i> , for SSDI-only beneficiaries who volunteer for BOND?	T22 to T21

In addition to the benefit offset, WIC and EWIC, and the change to an annual accounting period, some administrative differences might influence impact estimates for the first two research questions in the early years of BOND (including 2012). The administrative procedures established to provide T21 and T22 subjects with information and to implement benefit adjustments under the offset likely affected the speed with which retroactive payment adjustments were made and improper past payments recovered.

⁸ This percentage reflects two factors. First, of the beneficiaries solicited to volunteer for Stage 2, 53 percent came from the short-duration subpopulation, oversampling by a factor of 1.68. Short-duration beneficiaries were also more likely to volunteer once solicited: 6.4 percent did so compared to 4.2 percent of long-duration beneficiaries.

Because of how they are measured, these adjustments are especially important for the estimated impacts on benefits paid. By necessity, the impact estimates in this document focus on benefits paid *in* 2012. Impacts on benefits paid *for* 2012, which are not observed in the data available for this report, might be quite different, after all retroactive benefit adjustments and repayments of improper payments have been completed. For future reports, we plan to include estimates of the impact of BOND on benefits paid *for* the years in the evaluation period.

Finally, for the T21 to C2 comparison, T21 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. Although not intended, it is possible that in their implementation the two sets of counseling services differ in ways that have an impact on earnings and benefits above and beyond the impact of the offset itself.

1.3. Purpose

This *Snapshot Report* presents estimates for the three Stage 2 pairwise impact comparisons in the earliest years of the demonstration.⁹ The report refers to differences in T21 vs. C2 outcomes as *benefit offset impacts*, to T22 vs. C2 differences as *benefit offset plus EWIC impacts*, and to T22 vs. T21 differences as *EWIC vs. WIC impacts*. Because most subjects were not enrolled until 2012, this report focuses on impacts on earnings and benefit outcomes in 2012. A limited set of 2011 impacts appears in Appendix A.¹⁰

This first *Stage 2 Snapshot Report* uses the evaluation framework described in the *Evaluation Analysis Plan* (Bell et al. 2011).¹¹ Within that framework, the two most important evaluation outcomes—referred to as *confirmatory outcomes*—are total earnings and total SSDI benefits paid. In keeping with those designations, impacts on mean earnings in 2012 and mean benefits paid *in* 2012 serve as the confirmatory findings in this report. Hence, statistically significant findings for the confirmatory outcomes in this report should be interpreted as confirming that the benefit offset had an impact on at least one of two outcomes: 2012 earnings and/or SSDI benefits paid *in* 2012. The final impact evaluation will use a measure of benefits paid *for* the years in the evaluation period as a confirmatory outcome—the measure of benefits paid that is most important for policy purposes (but for which data are not currently available).¹²

⁹ Previous reports described the BOND design, the framework for estimating the impacts, and the early Stage 2 implementation activities (Stapleton et al. 2010; Bell et al. 2011; Gubits et al. 2013). A series of parallel reports documents results for Stage 1 of the demonstration.

¹⁰ Impacts on 2011 outcomes receive only limited attention in this report because only 40 percent of Stage 2 subjects had entered the demonstration by the end of 2011.

¹¹ Although the evaluation framework is the same as that described in the *Evaluation Analysis Plan*, some details of the estimation method have evolved from those described in the *Plan*. Appendix B of the current report describes the technical details of the estimation method used here to calculate the impacts presented in Chapter 3 and indicates which specific elements differ from methods in the *Plan*. Reasons for these deviations are also provided.

¹² Bell et al. (2011) identify benefits paid as the confirmatory measure and indicate only the difference between benefits paid *in* a period versus *for* a period in a footnote (footnote 40). It became apparent more recently during the preparation of the second Stage 1 snapshot report (Stapleton et al., 2014) that the difference between these two measures might be quite large, especially in the early demonstration years.

The report also presents exploratory impact findings for other beneficiary outcomes related to 2012 earnings and benefits paid in 2012. Significant findings for these outcomes cannot *confirm* that the benefit offset or EWIC had impacts; they can only suggest where such effects might have occurred. These estimates provide more information on the potential impacts of the benefit offset and EWIC, but receive less weight than the confirmatory findings in assessing the overall success of the tested treatments.

1.4. Organization of the Report

The remainder of this report consists of three chapters. Chapter 2 provides background information on the impact estimation methodology and descriptive findings that provide context for the impact estimates. Chapter 3 presents the impact findings for the confirmatory and exploratory outcomes. Chapter 4 includes a brief discussion of the results and their implications.

2. Methodology and Context

The goals for the Stage 2 evaluation are to learn about the impacts of the benefit offset for those most likely to use it (recruited and informed volunteers) and to determine the extent to which significant enhancements to the basic BOND-focused work incentives counseling affect offset utilization and impacts. For practical reasons, the design restricted the beneficiaries in Stage 2 to those most likely to use the offset. Specifically, attainment of the Stage 2 objectives requires more intensive data collection and more complex service delivery than is required for Stage 1. Restricting Stage 2 eligibility to those most likely to use the benefit offset reduces the sample sizes required for Stage 2 groups from tens of thousands to thousands.

Two aspects of this strategy for selecting the sample ensured that Stage 2 subjects would be likely to use the offset. First, concurrent beneficiaries were excluded from Stage 2. The interaction between SSI and SSDI substantially diminishes the value of the SSDI offset to concurrent beneficiaries, so it was expected that relatively few would use the SSDI offset. Second, in contrast to the Stage 1 sample (which is randomly selected from all eligible SSDI beneficiaries), the Stage 2 sample is composed of self-selected volunteers from randomly selected eligible SSDI-only beneficiaries. It is presumed that interest in using the offset led to the decision to volunteer for the study, and that this interest means that Stage 2 subjects will be more likely to use the offset than the average Stage 1 subject.¹³

For this report, administrative data for calculating earnings and benefit impacts were available through calendar year 2012. Earnings are measured from the SSA Master Earnings File (MEF), which contains longitudinal information on wages and self-employment income reported to the IRS. The MEF records were almost 100 percent complete for calendar year 2012 when SSA extracted them for this report. Benefit outcomes are measured from SSA's Payment History Update System (PHUS) for SSDI and the Supplemental Security Record (SSR), for SSI.¹⁴

The remainder of this chapter describes our methodological approach to estimating benefit offset impacts. We initially specified the methodology and outcomes for the impact analysis in Bell et al. (2011). This methodology was later refined for the *First-Year Stage 1 Snapshot Report* (Stapleton et al. 2013). For this report, we estimate impacts using a methodology that is largely similar to the Stage 1 methodology, but differs in that Stage 2 impacts are generalizable to the ten BOND sites only, rather than to the nation as a whole. We review the outcome definitions, anticipated impacts, estimation methodology, and analysis sample below.

¹³ A comparison of 2011 employment rates between the Stage 1 and Stage 2 samples shows that about 16 percent of Stage 1 subjects had at least some earnings in 2011, compared to about 37 percent of Stage 2 subjects.

¹⁴ Because the data are collected by the IRS and are therefore subject to IRS access rules, SSA staff have direct access to MEF data, but contractors do not. Consequently, qualified SSA staff accessed the data, submitted programs developed by the BOND Evaluation Team to estimate impacts, reviewed output to ensure that it complied with privacy requirements, and then transmitted the output to the evaluation team. The MEF earnings data are updated annually. The 2012 earnings data for this report were extracted in April 2014.

2.1. Outcome Definitions and Theoretical Impacts

The nine outcomes for this report include two confirmatory outcomes (total earnings and total SSDI benefits paid in 2012) and seven exploratory outcomes (related to employment and benefits). The exploratory earnings outcomes include indicators for earnings in excess of each of three annual earnings thresholds defined by multiples of BYA (one, two, and three times BYA) and an indicator for employment during 2012 (defined as any earnings in 2012). The exploratory benefit outcomes include number of months with SSDI payments, total SSI benefits paid, and number of months with SSI payments¹⁵—each in 2012. Impacts on 2011 outcomes for the full sample are presented in Appendix A and should also be considered exploratory.

In the discussion that follows, we consider the expected direction of benefit offset impacts on these outcomes, abstracting from administrative factors that could themselves influence the impacts. We then turn to a discussion of administrative factors and their potential influence on impacts.

Although BOND was designed to test whether eliminating the SGA cash cliff and replacing it with the \$1 for \$2 offset ramp would increase return to work and earnings, and reduce beneficiary's reliance on SSDI benefits (Bell et al. 2011), the theoretical direction of impacts of the benefit offset on mean earnings and benefits is ambiguous (third column of Exhibit 2-1). As described in detail in Bell et al. (2011), this ambiguity arises because the incentives created by the benefit offset vary with what the beneficiary's earnings would be under current law. T21 and T22 subjects who would have had no earnings or earnings below BYA under current law are expected, on average, to have higher earnings and lower SSDI benefits under the benefit offset. Conversely, some T21 and T22 subjects who would have had earnings well above BYA under current law are expected to have lower mean earnings and higher mean SSDI benefits under the benefit offset.¹⁶ Positive impacts on the mean earnings for all beneficiaries require that positive impacts for those whose earnings would be less than BYA under current law are sufficiently large to offset possible negative impacts for those who would earn more than BYA under current law.

Similarly, the predicted impact on benefits depends on what the earnings of the beneficiary would have been under current law. For those with no earnings or earnings below BYA, the predicted impact is negative; if they earn more than BYA under the offset than they would under current law, their benefits will fall. Conversely, for those who would have had earnings above BYA under current law, benefits for many under the offset are expected to be higher because they will be eligible for a partial benefit rather than no benefit at all, 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 must exceed the increase in benefits paid to those who would earn more than BYA under current law.

¹⁵ Although eligibility criteria for Stage 2 required that beneficiaries not be receiving SSI benefits at the time eligibility was determined (in the first six months of 2011), Stage 2 subjects could potentially become SSI recipients (for example, after spending down their assets enough to meet the resource test). Therefore, SSI benefits are included as an outcome variable.

¹⁶ Empirically, there is evidence 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.

Theory does, however, predict the signs of the impacts for five of the seven exploratory outcomes. It predicts positive impacts on employment, on the percentage of beneficiaries with earnings above BYA, and on months with SSDI payments. Theory also predicts negative impacts on SSI benefits and months with SSI payments. These predictions can be verified by separately considering the impacts for those whose earnings would be below or above BYA under current law. As indicated earlier, for those who would have earnings below BYA under current law, theory predicts that the offset will increase both the percentage employed and the percentage of beneficiaries with earnings above BYA. Those who would have earnings above BYA under current law will have a stronger incentive to keep their earnings above BYA under the offset than they do under current law—even though some might work and earn less under the offset. It is not possible to predict the direction of impacts on the percentage with earnings well above BYA (for example, two and three times BYA); however, it is expected that some T21 and T22 subjects whose earnings would be well above BYA under current law will reduce their earnings in response to the benefit offset.

Theory also predicts that the impact on SSI benefits paid will be negative. The offset might have an impact on SSI payments to T21 and T22 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.¹⁷

2.2. Administrative Features of the Offset That Influence Impacts

The previous discussion abstracts from the administrative features of the benefit offset that were designed and implemented to facilitate use of the offset by T21 and T22 beneficiaries. As described in Bell et al. (2011), because these processes are necessarily different from current law processes, they are part of the T21 and T22 interventions being tested under BOND.

In the first years of BOND, the administrative factors most likely to affect outcomes concern the administrative processes leading to the adjustment of benefits—the special processes implemented for T21 and T22 subjects and the current processes that apply to C2 subjects. For T21 and T22 subjects, that process started shortly after their enrollment date, when they were informed of their random assignment status. Some of those eligible to use the offset informed the demonstration of their work activity as recommended and their benefits were eventually adjusted via an administrative process set up for that purpose. Others eligible to use the offset early did not contact the demonstration, however. Instead, SSA discovered their high earnings in its annual review of earnings reported to the IRS, and then initiated the process to adjust their benefits.

The benefits measures for this report are based on benefits paid *in* 2012, rather than benefits paid *for* 2012, which includes all future retroactive adjustments for 2012 benefits. These two measures will diverge according to the dollar value of retroactive adjustments made for 2012 benefits. Although this dollar value is not yet known, we know that there must be retroactive adjustments of some dollar amount for the treatment subjects who did not pro-actively inform SSA of earnings above BYA. The BOND administrative data show that 34 percent of T21 subjects and 24 percent of T22 subjects eligible to use the

¹⁷ See Riley and Rupp (2012).

offset during 2011 and 2012 did not have a benefit adjustment until after 2012. This implies that some adjustments to benefits paid to T21 and T22 subjects for 2012 are not reflected in benefits paid *in* 2012, and that there will be at least some discrepancy between benefits paid *in* 2012 and benefits paid *for* 2012.

The direction and size of the impacts of this administrative factor depend on how the processes for the T21 and T22 groups compare to the corresponding processes for C2 subjects. The most striking difference is that T21 and T22 subjects had to be notified about a change in the earnings rules before the benefit adjustment process could start, whereas C2 subjects were subject to rules that had been in place for many years. Also, T21 and T22 administrative processes had not been previously implemented in a large scale, resulting in start-up delays¹⁸, whereas the C1 processes have been in place for many years.

Exhibit 2-1. Definitions of Confirmatory and Exploratory Outcomes and Predicted Signs of Impacts

	Definition	Predicted Sign
Confirmatory Outcomes		
Total earnings in 2012	2012 earnings	?
Total SSDI benefits paid in 2012	Sum of SSDI benefit payments from January through December 2012; for SSDI workers, this includes benefits for dependent spouses and minor children, but not for DAC ^a ; for DAC and DWB, it includes only benefits payable to the DAC or DWB	?
Exploratory Outcomes		
Earnings Outcomes (January–December 2012)^b		
Employment in 2012	Indicator for any 2012 earnings	+
Earnings above BYA	Indicator for 2012 earnings greater than or equal to \$12,120 (non-blind subjects) or \$20,280 (blind subjects)	+
Earnings above 2 × BYA	Indicator for 2012 earnings greater than or equal to \$24,240 (non-blind subjects) or \$40,560 (blind subjects)	?
Earnings above 3 × BYA	Indicator for 2012 earnings greater than or equal to \$36,360 (non-blind subjects) or \$60,840 (blind subjects)	?
Benefit Outcomes (January–December 2012)		
Number of months with SSDI payments	Number of months with SSDI benefits paid above zero	+
Total SSI benefits paid	Sum of SSI benefit payment amounts from January through December 2012	–
Number of months with SSI payments	Number of months with SSI benefits paid above zero	–

Notes: Bell et al. (2011) provide detailed discussion on the hypothesized impacts of benefit offset. The 2011 outcomes shown in Appendix A are defined in the same manner as the 2012 outcomes, with the sole difference being the time period of the measures.

^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 is based on earnings reported in the MEF.

¹⁸ This issue is described in Gubits et al. (2013) and Derr et al. (forthcoming).

One other administrative factor seems likely to have a positive impact on benefits paid *for* 2012, and possibly on benefits paid *in* 2012, but an ambiguous impact on 2012 earnings: the change from monthly to annual accounting. The purpose of this change was to simplify administration of the offset and to simulate the expected future accounting procedure should the benefit offset become national policy. While not the purpose of this change, the move to an annual accounting period is expected to help beneficiaries with highly variable earnings (for example, seasonal workers) to a significant degree. Under monthly accounting, earnings above SGA in any month reduce benefits for that month, but under annual accounting the benefit reduction for those same earnings might be smaller or zero because of earnings below the SGA amount in other months of the same year. Holding earnings constant, this administrative change is expected to increase the benefits paid to some beneficiaries; any increase in earnings due to this factor will reduce benefits (and correspondingly, any decrease in earnings will increase benefits). The theoretical sign of the impact of this administrative change on earnings is ambiguous.

2.3. Impact Estimation Methodology

SSA included Stage 2 in the demonstration in order to provide information about the impact of the benefit offset on beneficiaries who volunteer for the study and about the impact of EWIC vs. WIC. Given the self-selected nature of the Stage 2 sample, the impacts from Stage 2 should not be generalized to the national SSDI caseload or to any easily identifiable subpopulation. Instead, the Stage 2 impacts presented in this report generalize only to those who would have volunteered in the ten BOND sites had they been solicited.¹⁹

To estimate impacts, we compare mean outcomes for the T21, T22, and C2 groups to each other. The mean outcomes are weighted for differences in sampling rates into the solicitation pool across sampling strata and adjusted for the effects of small random differences in baseline characteristics.²⁰ The adjustments for differences in baseline characteristics also serve to reduce the standard errors. For each specific outcome, we test the null hypothesis of no impact. Each individual test uses a specified level of significance. For example, a 10 percent significance level means that if the null hypothesis is true, there is only a 10 percent chance that the test will mistakenly reject it.

The impact estimates are “intent to treat” estimates. For example, the benefit offset impacts capture the mean impact of the applicability of the benefit offset rules to the earnings of all T21 subjects, whether or not those subjects work and use the offset. Likewise, the benefit offset plus EWIC impacts capture the impact on all T22 subjects, whether or not they work. Hence, the impact estimates reflect “no impacts” for those treatment subjects who would not have any earnings under current law or the offset.

The Stage 2 impact analysis has a total of six confirmatory hypothesis tests: tests of impacts on the two confirmatory outcomes in each of the three pairwise comparisons. We group the four tests in the T21 vs. C2 and T22 vs. C2 comparisons together because they both involve impacts of the benefit offset. We perform a multiple comparison procedure on these four tests together to adjust the p-values of the tests. We perform a separate multiple comparison procedure to adjust the p-values of the two confirmatory tests in the T22 vs. T21 comparison. These adjustments are necessary because we are performing multiple

¹⁹ Notably, the findings do not generalize to all would-be volunteers in the service areas of all 53 SSA area offices in the nation.

²⁰ See Appendix B for a full description of the estimation model and the construction of analysis weights.

hypothesis tests, which makes the probability of at least one Type I error (rejecting a true null hypothesis) larger than the significance level for the individual tests. To compensate for this effect, we adjust the test statistics for the confirmatory tests so that the probability of rejecting the null hypothesis of no impact within the “family” of tests (i.e., either within the four tests of T21 vs. C2 and T22 vs. C2 or within the two tests of T22 vs. T21) is equal to the specified significance level if the null hypothesis of no impact on any outcome in the tested group is true.²¹

We make no multiple comparison adjustment to the tests for exploratory outcomes. Readers are advised to give less evidentiary weight to any individual significant result from an exploratory test than they would to an equally significant result from a confirmatory test.

We estimate impacts for the full Stage 2 assignment groups and for two pairs of subgroups, one defined by duration of SSDI benefit receipt at the point of solicitation into the demonstration²² and the other by year of enrollment into the study. The duration subgroups are of interest because prior research and program rules suggest that subjects who have been on the rolls for a *short duration* (defined here as three years or less) may respond to the benefit offset differently from those who have been on the rolls for a *long duration* (more than three years). More specifically, we expect more short-duration subjects to work in comparison to long-duration subjects. However, we expect it will take longer for short-duration subjects to actually have their benefits adjusted, because they will have completed fewer TWP and GP months at the outset of the demonstration in comparison to long-duration subjects.

The year of study enrollment subgroups are of interest for this report because of the limited length of follow-up offered by the 2012 outcomes. The 40 percent of T21 and T22 subjects who enrolled in the study during 2011 had at least a full calendar year to respond to their respective interventions, whereas those who enrolled during 2012 had less time to respond. Therefore, we look separately at these groups to explore whether there is evidence that longer exposure to the interventions increases the impacts of the interventions. In the future, with longer follow-up periods, we expect that any differences in impacts associated with the timing of study enrollment will recede in importance. We treat all subgroup analyses, including the tests of earnings and SSDI benefits paid, as exploratory.

2.4. Final Analysis Sample Sizes

Exhibit 2-2 presents the sizes for the overall sample and the subgroups. The final Stage 2 analysis sample contains a total of 12,744 subjects, spread across T21 (4,853), T22 (3,041), and C2 (4,850).

²¹ Our approach adjusts the *p*-values for the confirmatory outcomes using the Westfall-Young stepdown method. Details of the *p*-value adjustments for tests of impacts on the confirmatory outcomes appear in Appendix B. See Schochet (2009) for further discussion of the multiple comparisons problem.

²² We measure the duration of SSDI receipt from the outreach release date rather than from the date of random assignment in order to prevent endogenous selection into the duration subgroups. Some beneficiaries may have responded faster to outreach than others and the speed of their response may be correlated with their earnings and benefit outcomes. A short-duration beneficiary who took a long time to respond to outreach before enrolling in the study may have crossed the threshold into the long-duration definition (37 months or more of SSDI receipt) if duration is measured from random assignment. In order to rule out the possibility of subjects determining their subgroup membership after exposure to the study (which occurred when subjects were first solicited to enroll), we measure duration from outreach release date.

The baseline characteristics (not shown) for the T21, T22, and C2 samples are statistically equivalent to each other (Gubits et al. 2013). These descriptive findings give us a high level of confidence in the internal validity of the impact estimates. In other words, baseline equivalence bolsters the case that any study findings of statistically significant impacts represent real impacts of the interventions, rather than systematic preexisting differences between the three groups or their environments.

Exhibit 2-2. Stage 2 Analysis Sample Composition

Random Assignment Group	Full Sample	Duration		Year of Stage 2 Enrollment	
		Short Duration	Long Duration	2011	2012
Stage 2 Sample Unweighted Counts					
T21	4,853	3,124	1,729	1,948	2,905
T22	3,041	1,914	1,127	1,212	1,829
C2	4,850	3,103	1,747	1,941	2,909
Stage 2 Sample Weighted Percentages					
T21	100%	45.6%	54.4%	42.6%	57.4%
T22	100%	43.6%	56.4%	42.1%	57.9%
C2	100%	45.2%	54.8%	42.6%	57.4%

Source: BOND Operations Data System (BODS).

Notes: The Stage 2 analysis sample excludes 210 beneficiaries who are related to other BOND subjects (e.g., a primary and a DAC or two DACs with the same primary) to avoid contamination effects that might arise from the fact that almost all such beneficiaries (204 of the 210) were assigned to different BOND groups (see Appendix B for details on this adjustment). Because only six of these beneficiaries would have been able to be retained, it was not feasible to replicate the approach used for the Stage 1 analysis (where we were able to include pairs in which both members were assigned to the same group and revise the weights so that impact estimates reflect impacts for all beneficiary pairs with at least one member in Stage 1 (Stapleton et al. 2013)). Weights are used to account for differing probabilities of selection into the Solicitation Pool by site and duration of SSDI receipt. The weighted Stage 2 sample size is 49,763 (the estimated number of Stage 2-eligible beneficiaries in the ten BOND sites who would have volunteered had all Stage 2-eligible beneficiaries been offered the opportunity to enroll in the study).

3. Impact Findings

This chapter presents findings on the impact of Stage 2 of BOND in 2012, the first year in which all the beneficiaries who sought to qualify for the offset had entered the research sample. Those randomly assigned to one of the treatment groups became subject to the offset work incentives starting in 2011 (40 percent of T21 and T22 subjects) or during the first nine months of 2012 (60 percent). Hence, the duration of the demonstration's treatment was attenuated for the majority of sample members in the current reporting period, a factor that may limit measured impacts. Later reports will examine impacts in 2013 and beyond when all treatment group subjects are subject to the BOND benefit payment rules and work incentive counseling for an entire year or more.²³

For 2012 three policy comparisons are reported:

- The impact of the benefit offset with standard work incentives counseling (WIC) compared to current law (T21 vs. C2).
- The impact of the benefit offset and enhanced work incentives counseling (EWIC) compared to current law (T22 vs. C2).
- The incremental impact of adding EWIC to the benefit offset (T22 vs. T21).

For each policy comparison, we report estimates of impact on two confirmatory outcomes and seven exploratory outcomes, as shown in Exhibit 3-1. For each outcome, the table first presents regression-adjusted average outcomes for the three random assignment groups²⁴ and then supplies impact estimates reflective of the regression-adjusted differences between these mean outcomes. Thus for total earnings (first row of the exhibit), the estimated effect of the offset (plus WIC) compared to current law—shown in the fourth column as \$279—equals the difference between the average T21 outcome of \$3,929 and the average C2 outcome of \$3,650. Other impact columns and other rows of the exhibit follow this same structure.

As explained in Chapter 2, the significance levels for full-sample estimates of impacts on the confirmatory outcomes (total earnings and total SSDI benefits) are adjusted to address the multiple comparisons problem. The statistical significance of the confirmatory impact estimates at the 10-, 5-, and 1-percent significance levels are indicated with “#” symbols in the last three columns of the exhibit. For all other outcomes, and for all subgroup analyses, the impact estimates are considered exploratory and their significance levels are not adjusted for multiple comparisons. The significance levels of the exploratory estimates are indicated by asterisks. For the confirmatory outcomes, we describe estimates that are statistically significant at the 10-percent level as “some confirmatory evidence” of demonstration impact, while those significant at the 5-percent level are described as “confirmatory evidence” of impact

²³ Impacts in 2011 are reported in Appendix A of the current report, a period for which most treatment group subjects had no exposure to the BOND intervention and those that did so for just 4 months on average. We do not believe these results have meaning for policy but provide them for completeness and to support the later benefit-cost analysis (which will draw evidence of intervention impacts from the entire period of participation).

²⁴ The regression-adjusted average outcomes are calculated as the average predicted outcomes in the three groups using the common set of coefficients estimated in the regression model. See Section B.2 of Appendix B for a description of the regression model.

and those significant at the 1-percent level are characterized as “strong confirmatory evidence.” We term as “not statistically significant” any confirmatory impact estimate not significant at even the 10-percent level. Findings concerning exploratory outcomes are dubbed “suggestive” when found statistically significant at any of the three significance levels, since they are not adjusted to contain the heightened risk of false positive findings when multiple tests of significance are run.

3.1. Impacts for Full Stage 2 Sample

For the entire Stage 2 sample, we first discuss the confirmatory findings from Exhibit 3-1 and then consider exploratory evidence for the non-confirmatory outcomes. Later sections of the chapter break out findings into separate results for different subpopulations of the SSDI beneficiaries in the Stage 2 sample.

3.1.1. Confirmatory Impacts

We begin with the only confirmatory impact estimates in this report—total earnings and total SSDI benefits paid in 2012. Here we see some confirmatory evidence that the benefit offset increased both outcomes in 2012 relative to current law. The strongest findings are for earnings, which are very similar for the two types of counseling. Estimated impacts on earnings (first row of the exhibit) are \$279 for the comparison of offset-plus-WIC to current law and \$301 for the comparison of offset-plus-EWIC to current law. Both estimates equal about 8 percent of average earnings under current law (\$3,650) and together confirm that the presence of the offset makes a difference to beneficiary earnings compared to current law benefit provisions. However, there is no evidence that adding EWIC to the benefit offset had an added incremental impact: the T22 vs. T21 comparison provides a small (\$22) impact estimate that is not statistically significant.

There is also some confirmatory evidence that the offset increased total SSDI benefits paid in 2012 (sixth row of the exhibit). This is not unexpected given that some beneficiaries who would have gone into benefit suspension under current law continue to receive a partial benefit under the offset. While some of this automatic upward impact on benefits paid may have been countered by lower benefits paid to beneficiaries experiencing the earnings increases reported earlier, the net of these two influences was to slightly increase benefits on average for beneficiaries receiving the intervention. In particular, the estimated impact of the offset plus WIC in 2012 is \$148, or 1.2 percent of the mean under current law (\$12,558). It is also important to keep in mind that retroactive (post-2012) adjustments to the 2012 benefits of both treatment and control subjects will likely mean that the impacts on benefits paid *for* 2012 will ultimately be different than the impacts for benefits paid *in* 2012. The estimated impact for the offset plus EWIC compared to current law is somewhat smaller (\$109) and not statistically significant. Again, no effect is found for EWIC as an incremental addition to the offset; this estimate is small (– \$39) and not statistically significant.

3.1.2. Exploratory Impacts

The remaining rows of the exhibit provide suggestive evidence that the offset increased the proportion of sample members who had any employment during 2012 and the proportion with earnings above BYA. In the current law control group, 39.9 percent of beneficiaries had some employment in the year and 9.1 percent had earnings above the BYA. Findings suggest that the offset plus WIC increased the proportion employed by 1.6 percentage points and the proportion with earnings above the BYA by 1.3 percentage points. Both of these estimates are statistically significant without multiple comparison adjustments. The corresponding estimates from the offset-plus-EWIC to current law comparison are somewhat smaller in

magnitude and statistically significant only for the proportion with earnings above BYA. There is no evidence that EWIC does more than WIC to increase employment or the proportion with earnings above BYA.

**Exhibit 3-1. Estimated Impacts on 2012 Earnings and Benefits of Stage 2 Volunteers:
All Policy Comparisons**

	Average Outcome with Offset and WIC (T21) (1)	Average Outcome with Offset and EWIC (T22) (2)	Average Outcome under Current Law (C2) (3)	Estimated Impact of Offset + WIC vs. Current Law (T21 vs. C2) (4)	Estimated Impact of Offset + EWIC vs. Current Law (T22 vs. C2) (5)	Estimated Impact of EWIC Instead of WIC Given Offset (T22 vs. T21) (6)
Earnings Outcomes (January–December 2012)						
Total earnings (confirmatory)	\$3,929	\$3,951	\$3,650	\$279 # (\$128)	\$301 # (\$136)	\$22 (\$140)
Employment during year (%)	41.49	41.05	39.87	1.62 ** (0.81)	1.18 (0.90)	-0.44 (0.93)
Earnings above BYA (%)	10.37	10.10	9.11	1.26 ** (0.54)	0.99 * (0.59)	-0.27 (0.61)
Earnings above 2x BYA (%)	3.11	2.70	2.76	0.35 (0.32)	-0.06 (0.34)	-0.41 (0.34)
Earnings above 3 x BYA (%)	1.12	1.37	1.00	0.13 (0.20)	0.38 * (0.22)	0.25 (0.23)
Benefit Outcomes (January–December 2012)						
Total SSDI benefits paid (confirmatory)	\$12,705	\$12,664	12,558	\$147 # (\$65)	\$106 (\$72)	-\$41 (\$74)
Number of months with SSDI payments	11.42	11.47	11.38	0.04 (0.05)	0.09 * (0.05)	0.05 (0.05)
Total SSI benefits paid	\$50	\$50	\$40	\$10 (\$10)	\$10 (\$11)	-\$1 (\$12)
Number of months with SSI payments	0.18	0.18	0.16	0.02 (0.03)	0.02 (0.03)	0.00 (0.03)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the 10 study sites. Standard errors are in parentheses. Unweighted sample sizes: T21 = 4,853; T22 = 3,041; C2 = 4,850. See Exhibit 2-1 for variable definitions. Impact estimates are regression-adjusted for baseline characteristics. The Stage 2 subjects enrolled in the study from March 2011 through September 2012, with 40% of subjects enrolling in 2011 and 60% of subjects enrolling in 2012. The 2012 outcomes include at least some months prior to random assignment for the subjects who enrolled in the study during 2012. Total earnings and SSDI benefits paid are the two confirmatory outcome variables, and statistical tests for the impacts on these two outcomes used multiple comparison adjustments (see the Appendix for more details on the statistical tests and adjustments to the p-values). Tests for impacts on all other outcomes (exploratory outcomes) were conducted independently, without multiple comparison adjustments.

#/###/#### Impact estimate on confirmatory outcome is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test, after p-value has been adjusted by multiple comparisons procedure.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test (with no multiple comparisons adjustment).

Only 2.8 percent of beneficiaries had earnings above twice the BYA under current law, and only 1 percent had earnings above three times the BYA. The only suggestive evidence of impact for these indicators concerns the proportion of beneficiaries with earnings above three times the BYA, which is higher for the offset-plus-EWIC intervention compared to current law. Thus, although neoclassical economic theory predicts negative impacts for high earners, there is no suggestion of a negative impact on mean earnings for those with earnings above this relatively high level.

Most beneficiaries volunteering for the offset received SSDI payments in all or almost all months of 2012. The average number of months with SSDI payments was 11.4 under current law and slightly higher with the offset. As a result, we see suggestive evidence that the offset when combined with EWIC increased the number of months with benefits by a small amount (one tenth of a month on average).

The final two rows of the exhibit show that a small fraction of Stage 2 volunteers received SSI at some point in 2012, despite the fact that none of them did at the time they were solicited to volunteer for the demonstration. (Stage 2 excluded concurrent SSDI and SSI beneficiaries from its outreach.). Average yearly SSI benefits (including zeroes for the vast majority of beneficiaries who did not receive SSI) were around \$40 with no suggestive evidence that the tested interventions reduced their prevalence or amount.

3.2. Subgroups

We also considered how BOND affected the earnings and benefits of four subpopulations of beneficiaries who volunteered to receive the offset. These four subgroups are by short- and long-duration of SSDI participation and by year of demonstration entry, 2011 and 2012. Exhibits 3-2, 3-3, and 3-4 show impact estimates for subgroups defined by duration of SSDI receipt and Exhibits 3-5, 3-6, and 3-7 show estimates for subgroups defined by the year of enrollment into Stage 2. All subgroup analyses are exploratory. The significance tests are not adjusted for multiple comparisons. Therefore, at best, these subgroup results provide only suggestive evidence of impacts for subpopulations.

3.2.1. Subgroups Defined by Duration of SSDI Participation

We define short-duration beneficiaries as those who had received SSDI for up to three years (36 months) at the time they were solicited to volunteer for the study. All other sample members are considered long-duration beneficiaries. None of the differences between estimated impacts for the two subgroups (shown in column 7 of Exhibits 3-2, 3-3, and 3-4, which cover the three policy comparisons of interest) are statistically significant. Therefore, no additional insight beyond the full sample results is provided by the analysis of these two subgroups.

3.2.2. Subgroups Defined by Year of Demonstration Enrollment

Beneficiaries who entered the demonstration in 2011 have greater potential to have been affected by the offset and its work incentives counseling components during 2012 than beneficiaries who did not enroll in the demonstration until part way through the latter year. Exhibits 3-5, 3-6, and 3-7 split the sample on this basis, comparing estimated impacts on 2012 outcomes for 2011 enrollees (all of whom had been randomly assigned before January 2012) to those for 2012 enrollees (who were randomly assigned between January and September 2012 and thus had a shorter period in which they could have experienced impacts). None of the differences between estimated impacts on the two subgroups (shown in column 7 of the exhibits) are statistically significant. Therefore, no additional insight beyond the full sample results is provided by the analysis of these two subgroups.

Exhibit 3-2. Estimated Impacts of the Offset plus WIC Compared to Current Law (T21 Vs. C2) for Subgroups Defined by Duration of SSDI Receipt

	Short-Duration			Long-Duration			Estimated Difference in Impact (7)
	Average Outcome with Offset and WIC (T21) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and WIC (T21) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Earnings Outcomes (January–December 2012)							
Total earnings	\$3,910	\$3,513	\$396 ** (\$165)	\$3,944	\$3,763	\$181 (\$189)	\$215 (\$249)
Employment during year	39.51	38.17	1.34 (1.01)	43.13	41.27	1.85 (1.22)	-0.51 (1.58)
Earnings above BYA ^a	9.54	8.67	0.87 (0.60)	11.06	9.47	1.59 * (0.85)	-0.73 (1.04)
Earnings above 2x BYA	3.50	3.25	0.25 (0.37)	2.80	2.36	0.44 (0.50)	-0.19 (0.63)
Earnings above 3 x BYA	1.57	1.24	0.34 (0.26)	0.75	0.80	-0.05 (0.29)	0.39 (0.39)
Benefit Outcomes (January–December 2012)							
Total SSDI benefits paid	\$13,454	\$13,312	\$142 (\$87)	\$12,087	\$11,935	\$152 (\$93)	-\$11 (\$128)
Number of months with SSDI payments	11.56	11.51	0.05 (0.06)	11.31	11.28	0.03 (0.08)	0.02 (0.10)
Total SSI benefits paid	\$70	\$69	\$1 (\$17)	\$35	\$16	\$19 * (\$10)	-\$18 (\$20)
Number of months with SSI payments	0.23	0.22	0.01 (0.04)	0.13	0.11	0.02 (0.04)	-0.01 (0.05)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the 10 study sites. Standard errors are in parentheses. Unweighted sample sizes: short-duration: T21 = 3,124; short-duration C2 = 3,103; long-duration T21 = 1,729; long-duration C2 = 1,747. See Exhibit 2-1 for variable definitions. Impact estimates are regression-adjusted. The Stage 2 subjects enrolled in the study from March 2011 through September 2012, with 40% of subjects enrolling in 2011 and 60% of subjects enrolling in 2012. The 2012 outcomes include at least some months prior to random assignment for the subjects who enrolled in the study during 2012. Tests for impacts on all outcomes were conducted independently, without multiple comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test. †/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit 3-3. Estimated Impacts of the Offset plus EWIC Compared to Current Law (T22 Vs. C2) for Subgroups Defined by Duration of SSDI Receipt

	Short-Duration			Long-Duration			Difference in Impact (7)
	Average Outcome with Offset and EWIC (T22) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and EWIC (T22) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Earnings Outcomes (January–December 2012)							
Total earnings	\$4,054	\$3,513	\$541 *** (\$180)	\$3,871	\$3,763	\$108 (\$200)	\$433 (\$271)
Employment during year	39.34	38.17	1.17 (1.16)	42.47	41.27	1.20 (1.33)	-0.03 (1.77)
Earnings above BYA ^a	10.24	8.67	1.57 ** (0.70)	10.01	9.47	0.54 (0.91)	1.03 (1.16)
Earnings above 2x BYA	2.94	3.25	-0.31 (0.40)	2.50	2.36	0.14 (0.52)	-0.45 (0.66)
Earnings above 3 x BYA	1.92	1.24	0.68 ** (0.29)	0.93	0.80	0.13 (0.32)	0.55 (0.43)
Benefit Outcomes (January–December 2012)							
Total SSDI benefits paid	\$13,415	\$13,312	\$102 (\$83)	\$12,044	\$11,935	\$109 (\$112)	-\$7 (\$140)
Number of months with SSDI payments	11.59	11.51	0.08 * (0.05)	11.38	11.28	0.10 (0.08)	-0.02 (0.10)
Total SSI benefits paid	\$75	\$69	\$6 (\$20)	\$29	\$16	\$13 (\$12)	-\$7 (\$24)
Number of months with SSI payments	0.22	0.22	0.00 (0.04)	0.15	0.11	0.04 (0.04)	-0.04 (0.06)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the 10 study sites. Standard errors are in parentheses. Unweighted sample sizes: short-duration: T22 = 1,914; short-duration C2 = 3,103; long-duration T22 = 1,127; long-duration C2 = 1,747. See Exhibit 2-1 for variable definitions. Impact estimates are regression-adjusted. The Stage 2 subjects enrolled in the study from March 2011 through September 2012, with 40% of subjects enrolling in 2011 and 60% of subjects enrolling in 2012. The 2012 outcomes include at least some months prior to random assignment for the subjects who enrolled in the study during 2012. Tests for impacts on all outcomes were conducted independently, without multiple comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test. †/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit 3-4. Estimated Impacts of EWIC Compared to WIC, Given the Offset (T22 Vs. T21), for Subgroups Defined by Duration of SSDI Receipt

	Short-Duration			Long-Duration			Difference in Impact (7)
	Average Outcome with Offset and EWIC (T22) (1)	Average Outcome with Offset and WIC (T21) (2)	Impact Estimate (3)	Average Outcome with Offset and EWIC (T22) (4)	Average Outcome with Offset and WIC (T21) (5)	Impact Estimate (6)	
Earnings Outcomes (January–December 2012)							
Total earnings	\$4,054	\$3,910	\$145 (\$194)	\$3,871	\$3,944	-\$73 (\$197)	\$218 (\$274)
Employment during year	39.34	39.51	-0.17 (1.16)	42.47	43.13	-0.65 (1.40)	0.48 (1.82)
Earnings above BYA ^a	10.24	9.54	0.70 (0.71)	10.01	11.06	-1.05 (0.94)	1.75 (1.18)
Earnings above 2x BYA	2.94	3.50	-0.55 (0.40)	2.50	2.80	-0.30 (0.53)	-0.25 (0.66)
Earnings above 3 x BYA	1.92	1.57	0.35 (0.31)	0.93	0.75	0.18 (0.33)	0.17 (0.45)
Benefit Outcomes (January–December 2012)							
Total SSDI benefits paid	\$13,415	\$13,454	-\$39 (\$98)	\$12,044	\$12,087	-\$43 (\$108)	\$4 (\$145)
Number of months with SSDI payments	11.59	11.56	0.03 (0.07)	11.38	11.31	0.07 (0.08)	-0.04 (0.11)
Total SSI benefits paid	\$75	\$70	\$6 (\$22)	\$29	\$35	-\$6 (\$14)	\$11 (\$27)
Number of months with SSI payments	0.22	0.23	-0.01 (0.04)	0.15	0.13	0.01 (0.05)	-0.02 (0.06)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the 10 study sites. Standard errors are in parentheses. Unweighted sample sizes: short-duration: T22 = 1,914; short-duration T21 = 3,124; long-duration T22 = 1,127; long-duration T21 = 1,729. See Exhibit 2-1 for variable definitions. Impact estimates are regression-adjusted. The Stage 2 subjects enrolled in the study from March 2011 through September 2012, with 40% of subjects enrolling in 2011 and 60% of subjects enrolling in 2012. The 2012 outcomes include at least some months prior to random assignment for the subjects who enrolled in the study during 2012. Tests for impacts on all outcomes were conducted independently, without multiple comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test. †/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit 3-5. Estimated Impacts of the Offset plus WIC Compared to Current Law (T21 Vs. C2) for Subgroups Defined by Year of Study Enrollment

	2011 Enrollees			2012 Enrollees			Difference in Impact (7)
	Average Outcome with Offset and WIC (T21) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and WIC (T21) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Earnings Outcomes (January–December 2012)							
Total earnings	\$3,854	\$3,604	\$250 (\$212)	\$3,912	\$3,684	\$228 * (\$135)	\$22 (\$250)
Employment during year	41.57	38.99	2.58 * (1.32)	41.31	40.52	0.79 (0.99)	1.79 (1.64)
Earnings above BYA ^a	10.42	8.83	1.59 * (0.90)	10.19	9.31	0.88 (0.66)	0.71 (1.12)
Earnings above 2x BYA	3.22	3.02	0.21 (0.56)	2.92	2.57	0.34 (0.35)	-0.14 (0.68)
Earnings above 3 x BYA	1.10	0.94	0.17 (0.31)	1.06	1.04	0.02 (0.24)	0.15 (0.39)
Benefit Outcomes (January–December 2012)							
Total SSDI benefits paid	\$12,439	\$12,268	\$170 (\$108)	\$12,914	\$12,773	\$141 * (\$78)	\$29 (\$133)
Number of months with SSDI payments	11.30	11.25	0.06 (0.09)	11.51	11.48	0.03 (0.05)	0.03 (0.11)
Total SSI benefits paid	\$49	\$49	\$0 (\$15)	\$52	\$33	\$18 (\$13)	-\$19 (\$20)
Number of months with SSI payments	0.20	0.20	0.00 (0.05)	0.16	0.13	0.03 (0.03)	-0.03 (0.06)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the 10 study sites. Standard errors are in parentheses. Unweighted sample sizes: 2011 enrollees T21 = 1,948; 2011 enrollees C2 = 1,941; 2012 enrollees T21 = 2,905; 2012 enrollees C2 = 2,909. See Exhibit 2-1 for variable definitions. Impact estimates are regression-adjusted. The Stage 2 subjects enrolled in the study from March 2011 through September 2012, with 40% of subjects enrolling in 2011 and 60% of subjects enrolling in 2012. The 2012 outcomes include at least some months prior to random assignment for the subjects who enrolled in the study during 2012. Tests for impacts on all outcomes were conducted independently, without multiple comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test. †/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit 3-6. Estimated Impacts of the Offset plus EWIC Compared to Current Law (T22 Vs. C2) for Subgroups Defined by Year of Study Enrollment

	2011 Enrollees			2012 Enrollees			Difference in Impact (7)
	Average Outcome with Offset and EWIC (T22) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and EWIC (T22) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Earnings Outcomes (January–December 2012)							
Total earnings	\$3,936	\$3,604	\$332 (\$222)	\$3,738	\$3,684	\$54 (\$146)	\$278 (\$267)
Employment during year	40.12	38.99	1.13 (1.44)	41.35	40.52	0.83 (1.13)	0.30 (1.83)
Earnings above BYA ^a	10.14	8.83	1.31 (0.97)	9.64	9.31	0.32 (0.72)	0.99 (1.22)
Earnings above 2x BYA	2.89	3.02	-0.13 (0.59)	2.20	2.57	-0.37 (0.36)	0.25 (0.70)
Earnings above 3 x BYA	1.35	0.94	0.41 (0.34)	1.15	1.04	0.11 (0.26)	0.31 (0.43)
Benefit Outcomes (January–December 2012)							
Total SSDI benefits paid	\$12,537	\$12,268	\$269 ** (\$129)	\$12,793	\$12,773	\$28 (\$80)	\$248 (\$153)
Number of months with SSDI payments	11.41	11.25	0.16 * (0.09)	11.52	11.48	0.04 (0.06)	0.12 (0.11)
Total SSI benefits paid	\$51	\$49	\$2 (\$18)	\$49	\$33	\$16 (\$14)	-\$14 (\$23)
Number of months with SSI payments	0.23	0.20	0.03 (0.05)	0.15	0.13	0.02 (0.04)	0.01 (0.06)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the 10 study sites. Standard errors are in parentheses. Unweighted sample sizes: 2011 enrollees T22 = 1,212; 2011 enrollees C2 = 1,941; 2012 enrollees T22 = 1,829; 2012 enrollees C2 = 2,909. See Exhibit 2-1 for variable definitions. Impact estimates are regression-adjusted. The Stage 2 subjects enrolled in the study from March 2011 through September 2012, with 40% of subjects enrolling in 2011 and 60% of subjects enrolling in 2012. The 2012 outcomes include at least some months prior to random assignment for the subjects who enrolled in the study during 2012. Tests for impacts on all outcomes were conducted independently, without multiple comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test. †/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit 3-7. Estimated Impacts of EWIC Compared to WIC, Given the Offset (T22 Vs. T21) , for Subgroups Defined by Year of Study Enrollment

	2011 Enrollees			2012 Enrollees			Difference in Impact (7)
	Average Outcome with Offset and EWIC (T22) (1)	Average Outcome with Offset and WIC (T21) (2)	Impact Estimate (3)	Average Outcome with Offset and EWIC (T22) (4)	Average Outcome with Offset and WIC (T21) (5)	Impact Estimate (6)	
Earnings Outcomes (January–December 2012)							
Total earnings	\$3,936	\$3,854	\$82 (\$221)	\$3,738	\$3,912	-\$174 (\$151)	\$256 (\$268)
Employment during year	40.12	41.57	-1.45 (1.52)	41.35	41.31	0.04 (1.15)	-1.49 (1.90)
Earnings above BYA ^a	10.14	10.42	-0.28 (0.98)	9.64	10.19	-0.55 (0.75)	0.28 (1.24)
Earnings above 2x BYA	2.89	3.22	-0.33 (0.57)	2.20	2.92	-0.72 * (0.38)	0.38 (0.69)
Earnings above 3 x BYA	1.35	1.10	0.25 (0.36)	1.15	1.06	0.09 (0.27)	0.16 (0.45)
Benefit Outcomes (January–December 2012)							
Total SSDI benefits paid	\$12,537	\$12,439	\$98 (\$127)	\$12,793	\$12,914	-\$121 (\$88)	\$219 (\$155)
Number of months with SSDI payments	11.41	11.30	0.10 (0.10)	11.52	11.51	0.01 (0.06)	0.09 (0.12)
Total SSI benefits paid	\$51	\$49	\$2 (\$20)	\$49	\$52	-\$3 (\$16)	\$5 (\$26)
Number of months with SSI payments	0.23	0.20	0.02 (0.06)	0.15	0.16	-0.01 (0.04)	0.03 (0.07)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the 10 study sites. Standard errors are in parentheses. Unweighted sample sizes: 2011 enrollees T22 = 1,212; 2011 enrollees T21 = 1,948; 2012 enrollees T22 = 1,829; 2012 enrollees T21 = 2,905. See Exhibit 2-1 for variable definitions. Impact estimates are regression-adjusted. The Stage 2 subjects enrolled in the study from March 2011 through September 2012, with 40% of subjects enrolling in 2011 and 60% of subjects enrolling in 2012. The 2012 outcomes include at least some months prior to random assignment for the subjects who enrolled in the study during 2012. Tests for impacts on all outcomes were conducted independently, without multiple comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test. †/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

4. Discussion

The findings in this report apply only to the Stage 2 sample. The Stage 2 sample is composed of the SSDI beneficiaries thought most likely to respond to the offset work incentives. Specifically, the Stage 2 sample is made up of volunteers who wished to have the offset rules applied to them and who did not initially receive SSI.

Furthermore, the findings in this report apply only to calendar years 2011 and 2012. Calendar year 2012 is the first calendar year in which the offset and its work incentives counseling components could have affected the earnings and disability benefits of all treatment subjects in Stage 2 of BOND. Even in 2012, 60 percent of the Stage 2 treatment subjects had the opportunity to use the offset for less than the full year. Hence, these findings only begin to show the full effects of BOND for the Stage 2 sample.

The impact estimates show that:

- There is some confirmatory evidence that the offset rules combined with standard work incentives counseling (WIC) *increased mean earnings and SSDI benefits paid* compared to current law earnings rules and counseling services. There is also suggestive evidence (i.e., evidence from non-confirmatory tests) of increases in the proportions of beneficiaries with any earnings and with earnings above the BOND Yearly Amount (BYA).
- There is some confirmatory evidence that the offset rules combined with enhanced work incentives counseling (EWIC) *increased mean earnings* when compared to current law. There is also suggestive evidence of increases in the proportions of beneficiaries with earnings above BYA and above 3 x BYA, as well as an increase in the number of months with SSDI payments.
- When combined with the offset, EWIC *did not have any detectable effects on 2012 earnings or on benefit outcomes* relative to WIC.

The evidence of the offset impacts, given either type of counseling, arises for both short-duration (36 months or less) and long-duration SSDI beneficiaries, and is not appreciably stronger for those beneficiaries subject to the offset for the full year of study (2011 demonstration enrollees) than for those entering the demonstration part way through 2012.²⁵

Later reports will explore how these consequences of the benefit offset and special counseling services evolve over the longer run for Stage 2 subjects. Additional impact analyses will also examine new outcome measures (once data from the Stage 2 beneficiary follow-up surveys are available) and consider more beneficiary subgroups of interest to SSA. Finally, future impact analyses will consider impacts on SSDI benefits paid *for* the evaluation period. This measure will reflect retroactive adjustments to benefits, and so will be particularly informative in understanding the effects of the T21 and T22 treatments.

²⁵ Analysis of impacts by month of demonstration enrollment, rather than by year, also fails to reveal evidence of greater impacts for those beneficiaries with longer exposure to the offset—though the relatively small sample sizes of the individual monthly enrollment cohorts makes discovery of any trends that might be occurring difficult.

References

- Bell, Stephen H., Daniel Gubits, David Stapleton, David Wittenburg, Michelle Derr, Arkadipta Ghosh, and Sara Ansell. "BOND Implementation and Evaluation: Evaluation Analysis Plan." Submitted to the Social Security Administration. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, 2011.
- 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. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, forthcoming.
- 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.
- Riley, Gerald R., and Kalman Rupp. "Expenditure Patterns Under the Four Major Public Cash Benefit and Health Insurance Programs for Working-Age Adults with Disabilities." *Journal of Disability Policy Studies*, 2012. Published online ahead of print as doi:10.1177/1044207312469828.
- Schochet, Peter Z. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Evaluation Review*, vol. 33, no. 6, 2009, pp. 539-567.
- Social Security Administration. *Annual Statistical Report on the Social Security Disability Insurance Program, 2012*. Baltimore, MD: SSA, 2013.
- 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. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, 2014.
- Stapleton, David C., Stephen H. Bell, David C. Wittenburg, Brian Sokol, and Debi McInnis. "BOND Implementation and Evaluation: BOND Final Design Report." Submitted to the Social Security Administration, Office of Program Development & Research. Cambridge, MA: Abt Associates, December 2010.
- Weathers III, Robert R. 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.
- Wittenburg, David, David Stapleton, Michelle Derr, Denise W. Hoffman, and David R. Mann. "BOND Stage 1 Early Assessment Report. Final Report Submitted to the Social Security Administration." Cambridge, MA: Abt Associates, May 2012.

Appendix A. Stage 2 Impact Estimates on 2011 Earnings and Benefit Outcomes

Appendix A presents impacts on earnings and benefit outcomes measured in 2011 for the full Stage 2 sample. We note that only a minority of sample members could have experienced impacts in 2011, since 60 percent of them were not randomly assigned until 2012.

All impact estimates in Exhibit A-1 are exploratory. Although some of the estimates are statistically significant, they should be viewed with caution since the significance tests are not adjusted for multiple comparisons. The effect of offset plus EWIC compared to current law impact on earnings (\$265) is statistically significant while the offset plus WIC impact (\$148) is not, and the difference between the two estimates (\$116) is also not statistically significant.

Although SSDI beneficiaries who concurrently received SSI benefits were excluded from Stage 2, a small fraction received SSI after random assignment. Average yearly SSI benefits (including zeroes for the vast majority that did not receive SSI) were significantly higher for both treatment groups than for the control group (\$292 for T21 and \$294 for T22 versus \$229 for C1). Because, if anything, we expect a negative impact on SSI and because we do not observe a similar finding for 2012, this result should be viewed with caution.

**Exhibit A-1. Estimated Impacts on 2011 Earnings and Benefits of Stage 2 Volunteers:
All Policy Comparisons**

	Average Outcome with Offset and WIC (T21) (1)	Average Outcome with Offset and EWIC (T22) (2)	Average Outcome under Current Law (C2) (3)	Estimated Impact of Offset + WIC vs. Current Law (T21 vs. C2) (4)	Estimated Impact of Offset + EWIC vs. Current Law (T22 vs. C2) (5)	Estimated Impact of EWIC Instead of WIC Given Offset (T22 vs. T21) (6)
Earnings Outcomes (January–December 2011)						
Total earnings	\$3,072	\$3,188	\$2,924	\$148 (\$92)	\$264 ** (\$106)	\$116 (\$109)
Employment during year (%)	37.73	36.97	37.19	0.54 (0.77)	-0.21 (0.85)	-0.75 (0.86)
Earnings above BYA (%)	7.15	7.29	7.19	-0.04 (0.44)	0.11 (0.49)	0.14 (0.49)
Earnings above 2x BYA (%)	1.71	2.54	1.90	-0.19 (0.26)	0.64 ** (0.29)	0.84 *** (0.29)
Earnings above 3 x BYA (%)	0.68	0.89	0.59	0.09 (0.15)	0.30 * (0.16)	0.21 (0.18)
Benefit Outcomes (January–December 2011)						
Total SSDI benefits paid	\$13,082	\$13,081	12,986	\$96 (\$93)	\$95 (\$98)	-\$1 (\$98)
Number of months with SSDI payments	11.20	11.27	11.23	-0.03 (0.04)	0.04 (0.04)	0.07 (0.04)
Total SSI benefits paid	\$292	\$294	\$229	\$63 ** (\$24)	\$65 ** (\$27)	\$2 (\$30)
Number of months with SSI payments	0.20	0.21	0.19	0.02 (0.02)	0.02 (0.02)	0.00 (0.02)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the 10 study sites. Standard errors are in parentheses. Unweighted sample sizes: T21 = 4,853; T22 = 3,041; C2 = 4,850. See Exhibit 2-1 for variable definitions. Impact estimates are regression-adjusted for baseline characteristics. The Stage 2 subjects enrolled in the study from March 2011 through September 2012, with 40% of subjects enrolling in 2011 and 60% of subjects enrolling in 2012. Therefore the time periods for all outcomes are prior to the random assignment of most Stage 2 subjects. Tests for impacts on all outcomes were conducted independently, without multiple comparison adjustments.

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

Appendix B. Detailed Summary of Methodological Approach

This appendix describes the methods used to estimate impacts. Specifically, we describe the estimation procedure, the multiple comparisons adjustment, the covariates included in the estimation model, the analysis sample and weights, and the analysis of impact magnitudes by enrollment duration. The estimation procedure used in this report differs in several ways from that described in the *Evaluation Analysis Plan* (Bell et al., 2011) and from the precedent established in the earlier Stage 1 *Snapshot Reports*. We note these changes at the outset in Section B.1.

B.1 Changes from the *Evaluation Analysis Plan* and from Stage 1 Impact Estimation

The most notable change of the estimation procedure from the *Evaluation Analysis Plan* involves the external population to whom the Stage 2 impact estimates are intended to generalize. The *Evaluation Analysis Plan* indicated that Stage 2 impacts would represent the national population of beneficiaries (specifically, those Stage 2-eligible beneficiaries who would have volunteered for Stage 2 had they been offered the opportunity to enroll in the study). The method employed in this report instead only attempts to make Stage 2 impacts represent potential volunteers among the Stage 2-eligible beneficiaries in the ten BOND sites (rather than in the entire nation). This change was made after a test of the planned method resulted in instances of widely varying estimated standard errors across the three pairwise comparisons (T21 vs. C2, T22 vs. C2, and T22 vs. T21) for several outcomes.

This instability appears to be related to the relatively small number of clusters (i.e., the ten BOND sites). Clustered standard errors have downward bias and high variability when the number of clusters is small. The bias can be approximately corrected, but the high variability is a more challenging threat to the validity of confidence intervals and hypothesis tests, which are the two most important applications of standard errors. Therefore, the estimation method used in this report does not cluster errors at the site level. Unlike when attempting to generalize to the nation as a whole, the standard errors that only attempt to represent the 10 sites vary in the expected manner, with standard errors for the T21 vs. C2 comparison being slightly smaller than the standard errors for the other two comparisons. The relative sample sizes of the impact analysis samples leads to an expectation that the standard error magnitudes would follow this pattern.

A second change from the *Evaluation Analysis Plan* and from the precedent set by the Stage 1 *Snapshot Reports* involves how the multiple comparisons procedure was performed. The *Evaluation Analysis Plan* indicated that the multiple comparisons procedure would be applied in the final report to the two confirmatory outcomes (total earnings and total SSDI benefits) separately for each of the three pairwise comparisons in Stage 2. However, in the Stage 1 *Snapshot Reports*, we noted the importance of applying the multiple comparisons procedure in each report, rather than only in the final BOND report. Doing so will allow the final report's findings to emerge from a consistent method. We also implemented a resampling procedure (the Westfall-Young stepdown method) to adjust the p-values of the confirmatory hypothesis tests in the Stage 1 reports.

This report adheres to the decision to apply a multiple comparisons procedure to findings in each report. However, upon reconsideration of the plan to separately adjust the confirmatory tests for each pairwise comparison, we have decided that a more conservative approach for the T21 vs. C2 and T22 vs. C2

comparisons is prudent. These two comparisons both compare the impact of the offer of the benefit offset to a common control group that does not have the opportunity of using the offset. Since this gives the benefit offset *two* chances to create the appearance of a statistically significant impact on key outcomes by chance alone, we address all four confirmatory tests of these pairwise comparisons in a single multiple comparisons procedure. The test of EWIC vs. WIC embodied in the T22 vs. T21 comparison is not a test of the benefit offset, and so it seems appropriate to address multiplicity of tests separately for this comparison. Therefore, we perform one Westfall-Young procedure to address multiple tests in the T21 vs. C2 and T22 vs. C2 comparisons, and a second Westfall-Young procedure to address multiple tests in the T22 vs. T21 comparison.

A third change, compared to the Stage 1 approach, involves a simplification of the estimation method that is feasible for the smaller Stage 2 sample sizes. This change is described below in Section B.2.

A fourth change, involving a deviation from the *Evaluation Analysis Plan*, is in the construction of analysis weights. This multifaceted change (related both to the first change above and to the implementation of outreach) is described below in Section B.5.1.

Finally, a fifth change, compared to the Stage 1 approach, involves how the issue of related BOND subjects has been handled. The relatively small sample sizes of the Stage 2 assignment groups compared to the Stage 1 groups preclude the use of the weighting method that has been implemented to address this issue in the Stage 1 analysis. This issue is described further in Section B.5.2.

B.2 Estimation Procedure

Our basic impact estimation model is:

$$(1) \quad y_{ij} = \beta_0 + \beta_1 T_{21ij} + \beta_2 T_{22ij} + X_{ij} \Phi + \varepsilon_{ij}$$

where y_{ij} is an outcome measure for beneficiary i in site j ($j = 1, 2, \dots, 10$),

T_{21ij} = an indicator of whether beneficiary i in site j has been randomized into the T21 group (= 1 if so, = 0 if in T22 or C2 groups),

T_{22ij} = an indicator of whether beneficiary i in site j has been randomized into the T22 group (= 1 if so, = 0 if in T21 or C2 groups),

X_{ij} = a vector of baseline characteristics (listed in Section B.3) for individual i in site j ,

β_0 = the model intercept,

β_1 = the overall impact of the T21 treatment (vs. no treatment for the C2 group),

β_2 = the overall impact of the T22 treatment (vs. no treatment for the C2 group),

Φ = a vector of coefficients, and

ε_{ij} = an error term that is independent between and within sites.

In this model, the incremental impact of the T22 treatment compared with the T21 treatment is the difference $\beta_2 - \beta_1$.

Our statistical inferences for Stage 2 are limited to the self-selected population of SSDI-only beneficiaries in the 10 study sites who would volunteer for BOND if given the opportunity. Unlike the Stage 1 analysis, we are not attempting to make inferences about a national population of beneficiaries. Therefore, our analysis weights (see Section B.4) do not reweight the sites to reflect their probabilities of selection for the study, and our estimated standard errors and significance tests treat the 10 sites as fixed. We estimate model (1) by weighted least squares regression, using the SURVEYREG procedure in SAS. The estimated standard errors are computed by the Taylor series linearization method (also known as “robust standard errors”) without clustering. (We use the ESTIMATE statement to manipulate the estimated variance-covariance matrix to obtain estimated standard errors for the estimates of β_1 , β_2 , and $\beta_2 - \beta_1$.)

A less important difference between the Stage 1 and Stage 2 estimation procedures is that some of the computation-saving complexities of the Stage 1 approach are unnecessary given Stage 2’s much smaller sample sizes. The Stage 1 approach required multiple steps (initial regression in the control group, calculation of residuals for all subjects, formation of random groups for data reduction, and regression of average residuals on treatment status). For Stage 2, we accomplish the same goals by simply estimating the model (1) in one step instead of performing multiple steps.

For subgroup analyses, we use the following extension of model (1):

$$(2) \quad y_{ij} = \beta_0 + \beta_1 T_{21ij} + \beta_2 T_{22ij} + \beta_3 S_{ij} + \beta_4 (T_{21ij} S_{ij}) + \beta_5 (T_{22ij} S_{ij}) + X_{ij} \Phi + \varepsilon_{ij}$$

where

$S_{ij} = 1$ or 0 depending on which of two possible subgroups beneficiary i in site j belongs to,

β_1 = the impact of the T21 treatment (vs. no treatment for the C2 group) for the subgroup with $S_{ij} = 0$,

β_2 = the impact of the T22 treatment (vs. no treatment for the C2 group) for the subgroup with $S_{ij} = 0$,

β_3 = the difference between the two subgroups’ expected outcomes in the absence of treatment,

β_4 = the difference between the two subgroups in the impacts of the T21 treatment,

β_5 = the difference between the two subgroups in the impacts of the T22 treatment,

and the rest of the notation is as defined above. In this model, for the subgroup with $S_{ij} = 0$, the incremental impact of T22 vs. T21 is the difference $\beta_2 - \beta_1$. For the subgroup with $S_{ij} = 1$, the impact of T21 vs. C2 is $\beta_1 + \beta_4$, the impact of T22 vs. C2 is $\beta_2 + \beta_5$, and the incremental impact of T22 vs. T21 is $\beta_2 + \beta_5 - \beta_1 - \beta_4$. The difference between the two subgroups is the incremental impact of T22 vs. T21 is $\beta_5 - \beta_4$. Similar to the estimation of model (1), we used the SURVEYREG procedure in SAS to compute weighted least squares estimates of all model parameters and used the ESTIMATE statement to manipulate the variance-covariance matrix to obtain estimated standard errors for parameter estimates and their sums and differences.

B.3 Multiple Comparisons

The BOND impact analysis involves running a large number of hypothesis tests, because we analyze multiple outcome measures and multiple subgroups. Performing multiple tests increases the risk of “false positives”—that is, of finding statistically significant impacts for some outcomes or some subgroups even if the true impacts of BOND are zero. This increase in risk is called the “multiple comparisons problem.”

The impact analysis takes two measures to address the multiple comparisons problem:

1. We separated the hypothesis tests into “confirmatory” and “exploratory” tests, as specified in the *Evaluation Analysis Plan* (Bell et al. 2011), prior to conducting the impact analysis. Only the two most important outcomes—total earnings and total SSDI benefits paid—are included in the confirmatory group, and in this report, only the impact estimates for those outcomes in 2012 are treated as confirmatory. All other impact estimates, including all estimates for subgroups and all 2011 impacts, are considered exploratory. Statistically significant findings from confirmatory analyses are interpreted as evidence that the benefit offset had impacts on earnings or SSDI benefits. In contrast, statistically significant findings from exploratory analyses are characterized as suggestive.
2. We implemented a multiple comparisons adjustment procedure for our two confirmatory outcomes. The procedure controls the “familywise error rate”—the probability of rejecting at least one null hypothesis in a family of hypothesis tests when all null hypotheses are true.

The *Evaluation Analysis Plan* stated that the multiple comparisons adjustment would treat the three pairwise comparisons (T21 vs. C2, T22 vs. C2, and T21 vs. T22) as *separate studies* (i.e., the adjustment would be performed independently for each comparison). We have reconsidered this position and now believe that the pairwise comparisons of T21 vs. C2 and T22 vs. C2 should be handled together in the multiple comparisons adjustment. These two comparisons each compare a treatment group that has been offered the benefit offset against a control group that has not been offered the offset. This gives the benefit offset *two* chances to create the appearance of a statistically significant impact on key outcomes by chance alone. To control the overall risk of “false positives” in this sense, we adjust p-values from these two pairwise comparisons through a single procedure that keeps the overall probability of a false positive across four tests (T21 vs. C2 earnings, T21 vs. C2 SSDI benefits, T22 vs. C2 earnings, and T22 vs. C2 SSDI benefits) at or below the nominal significance level (.10, .05, or .01). In contrast, we view the EWIC vs. WIC (T22 vs. T21) comparison as a wholly different test of the effectiveness of enhanced work incentive services, rather than a further test of the benefit offset, so we handle this comparison separately from the other two. Thus, confirmatory analyses of the T22 vs. T21 comparison involve adjustment for only two tests (earnings and SSDI benefits), not four.

This report, like the Stage 1 snapshot reports, uses the Westfall–Young permutation stepdown method to control the familywise error rate. We describe the method below, starting with the example of adjusting for four tests.

In notation, let

- A, B, C, D = the four impact estimates of interest (in this case, impacts on earnings and SSDI benefits for the T21 vs. C2 and T22 vs. C2 comparisons), and

- $p_A^{raw}, p_B^{raw}, p_C^{raw}, p_D^{raw} = p$ -values from individual t-tests of impact estimates. These are the four “raw,” unadjusted p -values from the four standard significance tests taken independently.

We can then place the impact estimates in the order of their raw p -values.

- IMPACT1, IMPACT2, IMPACT3, IMPACT4 = the impact estimates ordered by raw p -values, from lowest raw p -value (IMPACT1) to highest (IMPACT4).
- $p_{IMPACT1}^{raw}, p_{IMPACT2}^{raw}, p_{IMPACT3}^{raw}, p_{IMPACT4}^{raw} =$ raw p -values in order from smallest to largest.

We then form some large number R permutation replicates. (The procedures used for this report use 20,000 replicates.) With each replicate sample, we run impact regressions for the four impacts, producing four p -values.

We can then define the adjusted p -values as follows²⁶:

$$\begin{aligned}
 p_{IMPACT1}^{adj} &= \frac{\text{Number of replicates where } \min \{p_{IMPACT1}^{rep}, p_{IMPACT2}^{rep}, p_{IMPACT3}^{rep}, p_{IMPACT4}^{rep}\} \leq p_{IMPACT1}^{raw}}{R} \\
 p_{IMPACT2}^{adj} &= \max \left\{ p_{IMPACT1}^{adj}, \frac{\text{Number of replicates where } \min \{p_{IMPACT2}^{rep}, p_{IMPACT3}^{rep}, p_{IMPACT4}^{rep}\} \leq p_{IMPACT2}^{raw}}{R} \right\} \\
 p_{IMPACT3}^{adj} &= \max \left\{ p_{IMPACT2}^{adj}, \frac{\text{Number of replicates where } \min \{p_{IMPACT3}^{rep}, p_{IMPACT4}^{rep}\} \leq p_{IMPACT3}^{raw}}{R} \right\} \\
 p_{IMPACT4}^{adj} &= \max \left\{ p_{IMPACT3}^{adj}, \frac{\text{Number of replicates where } p_{IMPACT4}^{rep} \leq p_{IMPACT4}^{raw}}{R} \right\}
 \end{aligned}$$

where p_{IMPACT}^{rep} is the p -value for an impact in a particular replicate.

For the T22 vs. T21 comparison, the method is analogous, but there are only two p -values to adjust, with

$$\begin{aligned}
 p_{IMPACT1}^{adj} &= \frac{\text{Number of replicates where } \min \{p_{IMPACT1}^{rep}, p_{IMPACT2}^{rep}\} \leq p_{IMPACT1}^{raw}}{R} \\
 p_{IMPACT2}^{adj} &= \max \left\{ p_{IMPACT1}^{adj}, \frac{\text{Number of replicates where } p_{IMPACT2}^{rep} \leq p_{IMPACT2}^{raw}}{R} \right\}
 \end{aligned}$$

Our significance tests for confirmatory impact estimates compare the adjusted p -values with the thresholds of .10, .05, and .01. Exhibit B-1 shows both unadjusted and adjusted p -values for the confirmatory impact estimates in this report.

²⁶ Westfall, Peter H., Randall Tobias, and Russell D. Wolfinger. *Multiple Comparisons and Multiple Tests Using SAS*. Cary, NC: SAS Institute. (2011): 349-350.

Exhibit B-1. Stage 2 Impact Estimates on Confirmatory Outcomes Illustrating the Multiple Comparison Adjustment on *p*-values

Comparison	Confirmatory Outcome	Impact Estimate (1)	<i>p</i> -value (Unadjusted) (2)	<i>p</i> -value (Multiple Comparisons Adjustment) (3)
First Multiple Comparison Procedure (4 hypothesis tests)				
T21 vs. C2	Total earnings in 2012	\$279# (\$128)	0.030	0.096
T21 vs. C2	Total SSDI benefits paid in 2012	\$147# (\$65)	0.022	0.096
T22 vs. C2	Total earnings in 2012	\$301# (\$136)	0.027	0.096
T22 vs. C2	Total SSDI benefits paid in 2012	\$106 (\$72)	0.141	0.149
Second Multiple Comparison Procedure (2 hypothesis tests)				
T22 vs. T21	Total earnings in 2012	\$22 (\$140)	0.877	0.881
T22 vs. T21	Total SSDI benefits paid in 2012	-\$41 (\$74)	0.577	0.830

Source: Analysis of SSA administrative records from the MEF, BODS, and MBR.

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the 10 study sites. Standard errors are in parentheses. Unweighted sample sizes: T21 = 4,853; T22 = 3,041; C2 = 4,850. See Exhibit 2-1 for variable definitions. Impact estimates are regression-adjusted for baseline characteristics.

###/### Impact estimate on confirmatory outcome is significantly different from zero at the .10/.05/.01 levels, respectively, using a two-tailed t-test, after *p*-value has been adjusted by multiple comparisons procedure.

B.4 Covariates

The use of baseline covariates in impact estimation increases the precision of impact estimates if they explain sufficient chance variation in the outcomes to compensate for their use of degrees of freedom, as is usually the case for large samples like those used here. Their use is not needed to achieve unbiasedness or consistency of the estimates. Our estimation model includes a set of covariates derived from administrative data (mimicking the Stage 1 impact model), as well as additional covariates derived from the Stage 2 baseline survey.²⁷ We used a machine learning procedure designed to improve statistical precision to assist us in the choice of covariates derived from the Stage 2 baseline survey.

Exhibit B-2 shows the administrative-data covariates. This list is based on the Stage 1 model (Stapleton et al. 2013, Exhibit A-2), with a few modifications. We excluded covariates related to Stage 2 eligibility (since they are constant for the Stage 2 sample) and added three new covariates:

- Randomly assigned in 2012 (dummy)
- Interaction of 2011 earnings and randomly assigned in 2012

²⁷ As described in the Stage 2 Early Assessment Report (p. 9, footnote 15), baseline survey information is available for all but 85 Stage 2 subjects.

- Interaction of 2010 earnings and randomly assigned in 2012

Each of the three new covariates equals zero for the subgroup of sample members randomly assigned in 2011, and each is a true baseline covariate in the sense that its value could not be affected by the subject's assignment to the T21, T22, or C2 group. The first two new covariates allow earnings in 2011 to help predict outcomes in 2012 for those randomly assigned in 2012. The third new covariate effectively allows the coefficient on 2010 earnings (which was already included in the Stage 1 model) to vary between the two random-assignment subgroups—the motivation being that for the 2012 subgroup, our predictors include both 2010 and 2011 earnings, while for the 2011 subgroup, we use only 2010 earnings.

Exhibit B-2. Administrative-Data Covariates Included in Stage 2 Impact Regressions

Covariates (measured at baseline unless otherwise specified)
Age
Age (squared)
AIME (Average Indexed Monthly Earnings) as of May 2011
AIME (Average Indexed Monthly Earnings) as of May 2011 (squared)
AIME (Average Indexed Monthly Earnings) as of May 2011 are equal to zero
Any employment in 2010 ^a
County 2010 employment rate for people with a disability
County April 2011 unemployment rate
Dummy for missing 2010 unemployment rate and missing rural status
Dummy for missing employment rate for people with a disability
Earnings in 2010 ^a
Gender
Has a representative payee
Has SSDI start date on or after January 1, 2010 (very short-duration beneficiary)
Interaction of very short-duration x 2010 earnings ^a
Interaction of monthly benefit amount at baseline and AIME as of May 2011
Interaction of age and number of years receiving SSDI
Interaction of earnings in 2010 and randomly assigned in 2012 ^a
Interaction of earnings in 2011 and randomly assigned in 2012 ^b
Is a disabled adult child (DAC) beneficiary
Is a disabled widow(er) beneficiary (DWB)
Is a dually entitled DAC beneficiary
Is a dually entitled DWB
Monthly benefit amount (MBA) at baseline
Monthly benefit amount (MBA) at baseline is equal to zero
Number of years receiving SSDI
Number of years receiving SSDI (squared)
Primary impairment category:
Neoplasms
Mental disorders
Back or other musculoskeletal
Nervous system disorders
Circulatory system disorders
Genitourinary system disorders
Injuries
Respiratory
Severe visual impairments
Digestive system
Other impairments
Unknown impairments

Covariates (measured at baseline unless otherwise specified)
Randomly assigned in 2012 ^a
Receives written beneficiary notices in Spanish
Rural area dummy
Short-duration SSDI receipt (36 months or fewer)
Site dummies
SSI receipt dummy

^a Included in model for all earnings outcomes and total SSDI benefits only.

^b Included in model for 2012 earnings outcomes and 2012 total SSDI benefits only.

We also include additional covariates from the Stage 2 baseline survey that were listed in the *Evaluation Analysis Plan* (shown below in Exhibit B-3).

Exhibit B-3. Survey Covariates Listed in the *Evaluation Analysis Plan* and Included in Stage 2 Impact Regressions

Covariates
Marital status (married, widowed, divorced, separated, never married)
Cohabiting dummy
Education dummies (LT HS, HS/GED, Some college, 4yr college degree)
Child under age 18 in household
Race/ethnicity (African American, Hispanic, White, Asian, Other)
Working at baseline (baseline survey)
Lives in non-group residence (single family home, regular apartment, or mobile home)
Enrolled in school or taking classes
Full-time student
Engaged in volunteer work
Health dummies (excellent, very good, good, fair, poor)
Number of months worked in previous 3 years
Square of number of months worked in previous 3 years
Personal goals include getting a job (if not working), moving up in a job, or learning new job skills
Health limits in moderate activities "a lot"
Health limits climbing several flights of stairs "a lot"
Emotional well-being (composite scale)
Stayed overnight in a hospital in past 12 months
Needs the help of another to get around inside home
Needs the help of another to get around outside home

Finally, prior to the impact analysis, we pre-specified a machine learning procedure for selecting additional covariates to improve statistical precision. Starting with a list of 59 potential covariates from the baseline survey, we used the lasso²⁸ to select additional variables and interaction terms that help predict 2012 earnings (without using the randomly assigned treatment status). The covariates selected by this procedure are shown in Exhibit B-4.

²⁸ Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* (1996): 267-288.

Exhibit B-4. Additional Covariates Selected by Pre-Specified Machine Learning Procedure

Covariates
Earned \$12,000 or more in past year
Change in health during past year (much better, somewhat better, about the same, somewhat worse, or much worse)
Not in the labor force
Self-employed
Employed at a steady job (neither temporary nor seasonal)
Employed full-time (35 or more hours per week)
Employed at a job with health insurance benefits
Employed at a job with many benefits (health insurance, paid sick days and vacation, long- and short-term disability benefits, and retirement benefits)
Able to drive a car
Has access to a car, truck, or van
Perceived barriers to employment (composite scale)
Able to do the same type of work as was doing when first became limited in the kind or amount of work or other daily activities one could do
Stayed in hospital more than 30 days in past year
Body mass index 25 or higher
Emotional problems limited activities most or all of the time
Interaction of employed full-time and rural area
Interaction of employed full-time and 4-year college degree
Interaction of employed full-time and engaged in volunteer work
Interactions of employed full-time and health dummies
Interaction of employed full-time and self-employed
Interaction of employed full-time and job with health insurance
Interaction of employed full-time and job with many benefits
Interaction of employed full-time and able to do the same type of work as was doing when first became activity-limited
Interaction of employed full-time and access to a car, truck, or van
Interactions of change in health and earned \$12,000 or more in past year
Interaction of not in the labor force and short-duration SSDI receipt
Interaction of not in the labor force and very short-duration beneficiary
Interaction of not in the labor force and 2010 earnings ^a
Three-way interaction of not in the labor force, 2010 earnings, and very short-duration beneficiary ^a
Interaction of self-employed and county April 2011 unemployment rate
Interaction of self-employed and age
Interaction of self-employed and squared age
Interaction of self-employed and able to drive a car
Interactions of employed at a steady job and primary impairment category
Interactions of employed at a job with health insurance benefits and site dummies
Interaction of employed at a job with health insurance benefits and MBA at baseline
Interaction of employed at a job with health insurance benefits and very short-duration beneficiary
Interaction of employed at a job with health insurance benefits and 2010 earnings ^a
Three-way interaction of employed at a job with health insurance benefits, 2010 earnings, and very short-duration beneficiary ^a
Interaction of employed at a job with many benefits and county 2010 employment rate for people with a disability
Interaction of employed at a job with many benefits and dummy for missing employment rate for people with a disability
Interactions of employed at a job with many benefits and marital status
Interactions of employed at a job with many benefits and race/ethnicity
Interactions of employed at a job with many benefits and health dummies

Covariates
Interactions of employed at a job with many benefits and health limits in moderate activities “a lot”
Interaction of perceived barriers to employment and earned \$12,000 or more in past year
Interactions of primary impairment category and able to do the same type of work as was doing when first became activity-limited
Interactions of MBA and able to do the same type of work as was doing when first became activity-limited

^a Included in model for all earnings outcomes and total SSDI benefits only.

B.5 Analysis Weights and Sample Exclusions

B.5.1. Construction of Analysis Weights

This section describes the construction of analysis weights for the impact analysis presented in this report. The construction of these weights differs from the approach described in the *Evaluation Analysis Plan* in three ways, which are described below.

- 1) **Removal of site component from the analysis weight.** In the *Evaluation Analysis Plan*, the Evaluation Team envisioned that Stage 2 results would represent Stage 2–eligible beneficiaries in the nation who would have volunteered had they been offered the opportunity to enroll in the study. This representation was to be accomplished through (a) standard errors that allow for random site-level variation in impacts and (b) weighting the Stage 2 sample to account for the random selection of BOND study sites. Section B.1 of this appendix describes the reason for having the Stage 2 sample represent would-be volunteers within the 10 BOND sites only, rather than within the whole nation. In addition to changing the estimation method for standard errors, this change also necessitates the removal of the weight component that accounts for the random selection of study sites.
- 2) **Additional complexity of probability of selection to outreach.** At the time the *Evaluation Analysis Plan* was written, the Evaluation Team did not foresee that operational concerns would require that the Solicitation Pool be composed of beneficiaries from three distinct sample files (delivered by SSA in December 2010, April 2011, and June 2011). Some beneficiaries on the two later files did not have the opportunity to be selected for the earliest outreach waves. The weights for this report take this fact into account, and reflect the actual probability of random selection into the outreach effort.
- 3) **Removal of assignment group component from the analysis weight.** As a final component to the Stage 2 analysis weight, the *Evaluation Analysis Plan* included the ratio of all volunteers to the number of volunteers randomly assigned to the beneficiary’s assignment group. This component would have served to equalize the weighted sample sizes between the three assignment groups (essentially weighting up the T22 group to equal the sizes of the larger T21 and C2 groups). Upon further consideration, it was noted that this component was unnecessary because of the lack of variation for this component within assignment group. Since this component is unnecessary for the unbiasedness of impact estimates, it was decided to simplify the weights by removing the component. The removal of this component means that rather than *each* assignment group weighting up to the number of would-be volunteers in the 10 BOND sites, the *entire* Stage 2 sample weights up to the number of would-be volunteers in the 10 BOND sites.

As described in Chapter 1, the goal of having at least 50 percent of volunteers be short-duration volunteers was accomplished through oversampling short-duration volunteers into the outreach effort. The Stage 2 analysis weights account for this oversampling. Each Stage 2 sample member is assigned an analysis weight given by:

$$w_{ikj} = \left(\frac{1}{P_{ikj}} \right)$$

where:

- w_{ikj} denotes the Stage 2 analytical weight for a volunteer of category i , site k , and type j (short- or long-duration), where i is one of four categories:
 1 = beneficiary included in December 2010 sample file only,
 2 = beneficiary included in December 2010 and April 2011 sample files,
 3 = beneficiary first included in April 2011 sample file,
 4 = beneficiary first included in June 2011 sample file;
- P_{ikj} denotes the probability of random selection to any one of the outreach waves of the Stage 2 recruitment effort for a volunteer of category i , site k , and type j ;

The probabilities of selection into any outreach wave differ for the four categories of beneficiaries, and are based on probabilities for inclusion into the three sets of outreach waves: (a) the pilot waves (January to April 2011), (b) the June 2011 wave, and (c) all other later waves from July 2011 to May 2012. The probabilities for inclusion into each of these sets of waves are:

$$P_{kj}(PILOT) = \left(\frac{N_{kj \text{ PILOT}}}{N_{1kj} + N_{2kj}} \right)$$

$$P_{kj}(JUNEWAVE) = \left(\frac{N_{kj \text{ JUNEWAVE}}}{(N_{2kj} - N_{2kj \text{ PILOT}}) + N_{3kj}} \right)$$

$$P_{kj}(LATERWAVES) = \left(\frac{N_{kj \text{ LATERWAVES}}}{(N_{2kj} - N_{2kj \text{ PILOT}} - N_{2kj \text{ JUNEWAVE}}) + (N_{3kj} - N_{3kj \text{ JUNEWAVE}}) + N_{4kj}} \right)$$

where:

- $P_{kj}(\bullet)$ denotes the probability of being included in a particular set of outreach waves for a Stage 2 eligible beneficiary in site k and of type j ;
- $N_{kj \bullet}$ denotes the number of subjects in site k of type j who were selected into a particular set of outreach waves;
- N_{ikj} denotes the total number of Stage 2 eligible beneficiaries of category i , site k , and type j ; and

- N_{ikj} denotes the number of subjects of category i , site k , and type j who were randomly selected into a particular set of outreach waves.

Using the probabilities for inclusion into a particular set of outreach waves, we can then define the probability of inclusion into *any* outreach wave for the four categories of volunteers:

$$P_{1kj} = P_{kj}(\text{PILOT})$$

$$P_{2kj} = P_{kj}(\text{PILOT}) + (1 - P_{kj}(\text{PILOT}))P_{kj}(\text{JUNEWAVE}) + \\ (1 - P_{kj}(\text{PILOT}) - (1 - P_{kj}(\text{PILOT}))P_{kj}(\text{JUNEWAVE}))P_{kj}(\text{LATERWAVES})$$

$$P_{3kj} = P_{kj}(\text{JUNEWAVE}) + (1 - P_{kj}(\text{JUNEWAVE}))P_{kj}(\text{LATERWAVES})$$

$$P_{4kj} = P_{kj}(\text{LATERWAVES})$$

It is important to note that the Stage 2 analysis weights *do not* align the weighted totals of the Stage 2 volunteers to represent all beneficiaries in the outreach waves. Only a small percentage (5 percent) of those solicited volunteered for the study. The Stage 2 analysis weights instead weight up the Stage 2 volunteers to represent those Stage 2-eligible beneficiaries in the ten BOND sites who would have volunteered had they been given the opportunity to enroll in the study. It is also important to note that the analysis weights give weight to the sites in proportion to the number of would-be volunteers in each site (with the number of would-be volunteers implied by the number of short- and long-duration Stage 2-eligible beneficiaries and the volunteer rates of short- and long-duration subjects in the site).

B.5.2. Sample Exclusions

The first Stage 1 Snapshot Report Appendix (Stapleton et al. 2013, pp. 34–35) explains an issue of “contamination” that could occur when two or more BOND subjects were “related”—specifically, when they were on the same primary SSDI beneficiary record and in the same site at baseline. The Stage 1 analysis excluded any set of related subjects who were randomly assigned to different groups. This exclusion was made because, for example, the behavior of a subject assigned to C1 could be influenced (“contaminated”) by a related subject’s assignment to T1. The analysis sample did include pairs of related subjects who were both in T1 or both in C1, weighting up these pairs to compensate for the exclusion of the contaminated pairs. Trios and larger sets of related subjects were excluded regardless of their random assignments. (Only one trio was completely assigned to T1, so there was no reliable way to represent trios in the analysis.)

The behavior of Stage 2 volunteers could be contaminated by the random assignments and behaviors of related subjects in the Stage 1 or Stage 2 experiments. Two related BOND subjects would end up in the same Stage 2 random assignment group only if *all* of the following three conditions were met:

1. In Stage 1, both subjects were randomly assigned to the Stage 2 solicitation pool.
2. Both subjects volunteered for Stage 2.
3. In Stage 2, both subjects were randomly assigned to the same group (T21, T22, or C2).

Since the probability of meeting all three conditions was low, Stage 2 volunteers who were related to other BOND subjects were very likely to be contaminated. Among the 12,954 Stage 2 volunteers, 210 were related to other BOND subjects, and only 6 of these 210 volunteers were uncontaminated (i.e., their BOND relatives ended up in the same Stage 2 random assignment group).²⁹ None of the uncontaminated volunteers were assigned to T22. While the 6 uncontaminated subjects could have been included in the T21 vs. C2 comparison, they would have had to be dropped in the T22 vs. C2 and T22 vs. T21 comparisons as there were no related subject counterparts in T22. In addition, these 6 subjects would have had very large analysis weights in the T21 vs. C2 comparison in order to represent the related BOND subjects who were contaminated. We therefore decided to exclude all of the 210 volunteers with BOND relatives. The resulting Stage 2 analysis sample has 12,744 subjects. The sample does not represent volunteers with BOND-eligible relatives, but the excluded subset is only a small fraction of the original Stage 2 sample (1.6 percent unweighted; 1.8 percent using the analysis weights from Section B.4.1).

²⁹ Of the 210 beneficiaries related to other BOND subjects, 189 beneficiaries were related to one other BOND subject, 18 were related to two other subjects, 2 were related to three other subjects, and 1 was related to 4 other subjects. Of the 6 subjects who were uncontaminated, two pairs of related subjects were assigned to T21 and one pair of related subjects was assigned to C2.