BOND Implementation and Evaluation

Third-Year Snapshot of Earnings and Benefit Impacts for Stage 1

Deliverable 24.c.3

April 22, 2015

Submitted to:
Social Security Administration
Attn: Ms. Joyanne Cobb
Office of Research, Demonstrations, and Employment Support
500 E Street, SW
9th Floor, Room 90528
Washington, DC 20254

Contract No. SS00-10-60011

Prepared by:
David Wittenburg
David R. Mann
David Stapleton
Daniel Gubits
David Judkins
Andrew McGuirk

MATHEMATICA
CENTER FOR
STUDYING DISABILITY POLICY

Abt
ASSOCIATES

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 study impacts for those most likely to use the offset (recruited and informed volunteers) and to determine the extent to which significant enhancements to counseling services affect impacts.

This report, the third in a series of *Snapshot Reports,* documents Stage 1 impacts on earnings and benefit outcomes during the third calendar year of implementation (2013). As in the previous *Snapshot Reports* (Stapleton et al. 2013, 2014), we estimate impacts by comparing earnings and benefit outcomes for all Stage 1 treatment subjects (T1) with those for all control subjects (C1). Future annual reports will track Stage 1 impacts through 2017. A parallel series of *Snapshot Reports* are being produced for Stage 2.

Summary of Key Findings

The benefit offset, as administered under BOND, did not have a statistically significant impact on total earnings in 2013—similar to previously reported findings for 2011 and 2012. In addition, we find no evidence of an increase in the proportion earning above the BOND Yearly Amount (BYA)—the annualized substantial gainful activity amount.

As in 2012, there was a positive impact on total SSDI benefits paid *in* 2013, an impact estimated as twice the size of the 2012 estimate. As in previous reports, the impact on benefits paid *in* 2013 does not represent the impact on benefits that will be eventually paid *for* 2013, because the benefits paid *in* 2013 do not reflect retroactive adjustments to benefit payments made after 2013, including adjustments for the benefit offset. It is not possible with existing data to assess the size of impacts on benefits paid *for* 2013, but the absence of an impact on the percentage with earnings above BYA noted above suggests that the impact on benefits paid *for* 2013 will be positive. The offset can reduce benefits paid for a given year only by inducing beneficiaries to earn above BYA when they would not otherwise have done so.

Several factors suggest that positive impacts on earnings may yet emerge, including growth in the number of treatment subjects whose benefits have been adjusted under the offset. These factors do not guarantee positive impacts at any point, however. Future reports will continue to document annual earnings and benefit impacts.

The BOND Evaluation Team

Abt Associates, in partnership with 25 other organizations, is implementing and evaluating BOND under contract to the SSA. 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. Neither the co-directors nor any members of their evaluation team have any role in implementing or overseeing the BOND intervention they are studying. Separation of implementation and evaluation does not extend throughout the project, however. Project Director Michelle Wood and Principal Investigator Howard Rolston of Abt 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 rest with the evaluation co-directors.

Table of Contents

Abbreviations Used in This Report

AWI	Average Wage Index
BODS	BOND Operations Data System
BOND	Benefit Offset National Demonstration
BYA	BOND Yearly Amount
CPI	Consumer Price Index
DAC	Disabled Adult Child
DWB	Disabled Widow/Widowers
EIE	Earned Income Exclusion
EWIC	Enhanced Work Incentive Counseling
GP	Grace Period
IRS	Internal Revenue Service
IRWE	Impairment Related Work Expense
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 Incentive Counseling
WIPA	Work Incentives, Planning, and Assistance

1. Introduction

The Benefit Offset National Demonstration (BOND) is a random assignment demonstration that tests one variant of Social Security Disability Insurance (SSDI) program rules governing work and other supports. This *Snapshot Report* concerns Stage 1 of BOND, which was designed to examine how a national benefit offset and accompanying administrative changes would affect earnings and program outcomes for the entire SSDI population. Stage 2 was designed to learn more 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 in Stage 1. This report is the third in a series of Stage 1 *Snapshot Reports* about impacts on earnings and benefits paid. This introductory chapter provides a synopsis of the demonstration, then describes the purpose of the report and outlines its remainder. Impact estimates for Stage 2 will appear in a separate series of reports.

1.1. Synopsis of BOND

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 after completing a nine-month Trial Work Period (TWP) and a three-month Grace Period (GP), an SSDI beneficiary's countable monthly earnings exceed the monthly Substantial Gainful Activity (SGA) amount. In 2013, the SGA amount was $1,040 per month for non-blind beneficiaries and $1,740 per month for blind beneficiaries. The complete loss of benefits for earnings in excess of the SGA amount is sometimes called the "cash cliff." It likely discourages some beneficiaries from working at all and encourages those who could work with earnings above the SGA level to keep their earnings below that level.

BOND replaces the cash cliff with a "ramp" (benefit offset) and a policy objective of encouraging beneficiaries who can work above the SGA level to increase their earnings and reduce their reliance on benefits. More specifically, the benefit offset is expected to increase the earnings of those who might otherwise not work at all, or might not attempt to earn more than the SGA amount. If such individuals engage in SGA under the benefit offset, their benefits will ultimately be reduced. Offsetting the possible benefit reduction for this group are benefits paid under BOND to those who would have had earnings above the SGA amount in the absence of BOND. Thus, the direction of the net impact on mean earnings and benefits of all beneficiaries will depend on the size of the impacts for beneficiaries who would not engage in SGA under current law relative to the size of the impacts for those who would. Members of the latter group lose their benefits entirely under current law, whereas under the benefit offset, many (perhaps most) will be eligible for a reduced SSDI benefit. While still on the ramp—that is, 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 beneath their full earnings potential.

BOND also changes the administrative processes to adjust benefits, including replacing the monthly SGA calculation, with an annualized measure of SGA referred to as the BOND Yearly Amount (BYA). BYA is equal to 12 times the monthly SGA amount (in 2013, $12,480 for non-blind and $20,880 for blind T1 subjects). The benefit offset reduces benefits by $1 for every $2 in countable annual earnings in excess 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 also receive Supplemental Security Income, Medicaid, or a variety of other public or private benefits that are contingent on earnings in some fashion.

the BYA following the GP. 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 is based on expected annual earnings. In the following calendar year, SSA reconciles payments with actual countable earnings, based on information provided by the Internal Revenue Service (IRS), documentation provided by the beneficiary, or both.

T1 subjects also have access to counseling services that are tailored to the benefit offset but otherwise designed to be comparable to counseling services available to all beneficiaries under current law. Although this effect is not intended, the counseling services implemented may be having an impact on earnings and benefits above and beyond the impact of the offset itself.

By necessity, the impact estimates in this document focus on benefits paid *in* 2013 rather than benefits paid *for* 2013. The distinction between these two measures exists because benefits paid *in* 2013 do not reflect retroactive payment adjustments made and improper past payments recovered after 2013 for the 2013 calendar year. Impacts on benefits paid *for* 2013, which are not observed in the data available for this report, may be quite different than benefits paid in 2013, after all retroactive benefit adjustments and repayments of improper payments have been completed. In contrast, this administrative issue does not affect the earnings impact estimates in this report.

The evaluation team is responsible for all the estimates that appear in this report. In previous reports, we described the BOND design and the framework for estimating the impacts, summarized early assessment activities on the infrastructure to support Stage 1 service delivery, and reported 2011 and 2012 earnings and benefit impacts for Stage 1 subjects (Stapleton et al. 2010; Bell et al. 2011; Wittenburg et al. 2012; Stapleton et al. 2013; Stapleton et al. 2014, respectively). The team is documenting the outcomes of Stage 2 in a series of parallel reports.

1.2. Purpose

This *Snapshot Report* presents estimates of the combined impact of the benefit offset and accompanying administrative changes for Stage 1 (hereafter referred to as *benefit offset impacts*) from January through December 2013. Random assignment occurred in May 2011, so the 2013 calendar year represents the third year and the second full calendar year of the evaluation period.

This third *Snapshot Report* uses the same analysis sample as the previous *Snapshot Reports* (Stapleton et al. 2013, 2014). The only notable change is that we used a new method to compute standard errors to improve their stability. We detail the new standard error calculation method in Appendix A. Appendix B contains the outcome tables in the previous *Snapshot Reports* repopulated using updated standard errors. Despite the changes in method to calculate standard error, the magnitude of the impact estimates in Appendix B are the same as those shown in previous *Snapshot Reports*.

The two most important evaluation outcomes—referred to as *confirmatory outcomes*—are total earnings and total SSDI benefits paid. In keeping with that structure, mean earnings in 2013 and mean benefits paid *in* 2013 are the confirmatory outcomes in this report. However, the discussion in this report places greater emphasis on 2013 earnings because of the expectation that benefits paid *for* 2013, after retroactive adjustments and recovery of improper payments, will ultimately differ substantially from benefits paid *in* 2013. 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: 2013 earnings and/or

SSDI benefits paid *in* 2013. The findings do not provide information on SSDI benefits ultimately to be paid *for* 2013, because we cannot measure this outcome with the available administrative data for this report. The final impact evaluation will use as the confirmatory outcome a measure of benefits paid *for* the years in the evaluation period—the measure of benefits paid that is most important for policy purposes. The report also presents estimates for several exploratory outcomes related to 2013 earnings and benefits paid, to explore the possibility of other impacts of the offset. These findings do not confirm that the benefit offset had impacts; they only suggest where such effects might have occurred. In providing more information on the potential impacts of the benefit offset, these findings receive less weight than the confirmatory findings in assessing the overall success of the tested treatment.

1.3. Organization of the Report

The remainder of this report comprises 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.

2. Methodology and Context

The goal for the Stage 1 evaluation is to learn about offset utilization and key impacts when the benefit offset is offered to all SSDI beneficiaries. The evaluation for Stage 1 compares outcomes for beneficiaries who were randomly assigned to the following groups:[2]

- **T1 subjects**: beneficiaries whose benefits are determined by the benefit offset rules over a period of at least five years and who have the opportunity to use ancillary demonstration services.
- **C1 subjects**: a control group that continues to receive benefits according to current law.

Because only a small fraction of T1 subjects will likely use the benefit offset, the T1 and C1 groups must be very large (tens of thousands each) to detect policy-relevant impacts (Stapleton et al. 2010). The BOND sample for random assignment included all SSDI beneficiaries between the ages of 20 and 59 in 10 randomly selected sites throughout the nation who were receiving benefit payments in April 2011.

For this report, administrative data for calculating earnings and benefit impacts were available through calendar year 2013. 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 2013, 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).[3]

The next section of this chapter describes our approach to estimating benefit offset impacts. We then summarize findings related to benefit offset usage from our previous reports and descriptive data on offset adjustments that might influence the size and direction of earnings and benefits impacts presented in Chapter 3.

2.1. Methodological Approach

We initially specified the methodology and outcomes for the impact analysis in Bell et al. (2011), and later implemented with only minor revisions the methodology for 2011 impacts in Stapleton et al. (2013). For this report, we revised the approach to calculating standard errors for the impact estimates. We summarize the new procedure below and describe it in detail in Appendix A. All other analysis methods used, the analysis sample, and the outcome definitions in this report are the same as those used in the 2011 and 2012 *Snapshot Reports*. Below, we review the outcome definitions, anticipated impacts, estimation methodology (including revisions to the standard errors), and analysis sample.

---

[2] The random assignment process included beneficiaries for Stages 1 and 2. Beneficiaries were randomly assigned to the T1 group, the C1 group, or a "Stage 2 solicitation pool." The latter initially included all random replicates for the purposes of recruiting volunteers for Stage 2. When Stage 2 recruitment was complete, subjects in the unused random replicates were assigned to C1. For details on random assignment, see Bell et al. (2011).

[3] 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 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 November 2013.

2.1.1. Outcome Definitions and Theoretical Impacts

The nine outcomes for this report include two confirmatory outcomes (total earnings and total SSDI benefits paid in 2013) 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 (the BYA, and two and three times the BYA) and an indicator for any earnings during 2013. The exploratory benefit outcomes include number of months with SSDI payments, total SSI benefits paid, and number of months with SSI payments—each in 2013.

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 consider administrative factors and their potential influence on impacts.

Although BOND was designed to test whether replacing the SGA cash cliff 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 detailed in Bell et al. (2011), this ambiguity arises because the incentives created by the benefit offset vary with the beneficiary's expected earnings under current law. T1 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 T1 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.[4] 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 beneficiaries 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.

Theory does, however, predict the signs of the impacts for five of the seven exploratory outcomes. It predicts positive impacts on the proportion employed, the proportion with earnings above BYA, and months with SSDI payments and 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

[4] There is empirical 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.

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 T1 subjects whose earnings would be well above BYA under current law will reduce their earnings in response to the benefit offset.[5]

Theory also predicts that the impact on SSI benefits paid will be negative. Under current law, any beneficiary who concurrently receives SSDI and SSI (a "concurrent" beneficiary) and is engaged in SGA after completing the TWP and GP is entitled to, at most, only an SSI payment.[6] In contrast, a concurrent T1 subject with the same earnings would likely receive a partial SSDI benefit, and the size of the T1 subject's SSI benefit would be reduced by the amount of the partial SSDI benefit or by the entire current-law SSI payment if the latter is smaller than the partial SSDI benefit. The offset might also have an impact on SSI payments to T1 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.[7]

2.1.2. 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 T1 beneficiaries. Because these processes differ from current law processes, they are part of the T1 intervention being tested under BOND (Bell et al. 2011).

The administrative adjustment of benefits for BOND subjects—the special process implemented for T1 subjects in contrast to the current law process that applies to C1 subjects—has the potential to affect the measurement of impacts on SSDI benefits, but not on annual earnings. The data available for this report

---

5 The predicted effect of the offset is positive on the percentage of T1 subjects with earnings above BYA, but is theoretically ambiguous on the percentage with earnings above higher thresholds (for example, two and three times BYA). The variation in the direction of the predicted earnings response by initial earnings level is the reason that the sign of the predicted impact on mean earnings is ambiguous (see Bell et al. 2011 for more details).

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

7 See Riley and Rupp (2012).

do not reflect benefit adjustments for the 2013 calendar year that were not implemented until 2014 or later. As a result, impacts reported in this report are for benefits paid *in* 2013, not benefits paid *for* 2013. There will be some difference between the mean benefit outcomes reported here (those paid in 2013) and mean benefit outcomes after all the retroactive adjustments to benefits for 2013 are made, for both treatment and control subjects. The consequences of these adjustments for estimated impacts will depend on how adjustments for the T1 subjects compare with the adjustments for C1 subjects, and are currently unknown.[8]

One other administrative factor seems likely to have a positive impact on benefits paid *for* 2013, and possibly on benefits paid *in* 2013, but an ambiguous impact on 2013 earnings: the change from monthly to annual accounting. Although this change was intended to simplify administration of the offset, it is expected to also help to a significant degree beneficiaries with highly variable earnings (for example, seasonal workers). Under monthly accounting, earnings above SGA in any month reduce benefits for that month, but under annual accounting, the benefit reduction 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 (decrease) in earnings due to this factor will reduce (increase) benefits. The theoretical sign of the impact of this administrative change on earnings is ambiguous.

[8] See Stapleton et al. (2014) for descriptions of the differences in the adjustment processes for T1 and C1 subjects.

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

	Definition	Predicted Sign
Confirmatory Outcomes		
Total earnings in 2013	2013 earnings	?
Total SSDI benefits paid in 2013	Sum of SSDI benefit payments from January through December 2013; for SSDI workers, this amount 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 2013)[b]		
Employment in 2013	Any 2013 earnings	+
Earnings above BYA	2013 earnings above $12,120 (non-blind subjects) or $20,280 (blind subjects)	+
Earnings above 2 × BYA	2013 earnings above $24,240 (non-blind subjects) or $40,560 (blind subjects)	?
Earnings above 3 × BYA	2013 earnings above $36,360 (non-blind subjects) or $60,840 (blind subjects)	?
Benefit Outcomes (January–December 2013)		
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 2013	–
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.

[a] For a description of family benefits, see [http://www.socialsecurity.gov/pubs/10024.html#a0=3]; accessed January 9, 2015.

[b] Earnings relative to BYA is based on earnings reported in the MEF, without adjustment for impairment-related work expenses (IRWE). Less than 1 percent of SSDI and SSI beneficiaries use IRWEs (Livermore et al. 2009); even when used, IRWEs do not appear in administrative records until claimed by the beneficiary and approved by SSA.

2.1.3. Impact Estimation Methodology

The goal of the Stage 1 BOND experiment is to make inferences about what the impact of the benefit offset would be if it applied to all SSDI beneficiaries in the nation who met the BOND eligibility criteria as of May 2011. The statistical design for the demonstration supports the production of unbiased point estimates and standard errors for this population. The standard errors reflect random variation associated both with the selection of the BOND sites and with assignment of subjects in those sites to T1 and C1. As a result, each test of a null hypothesis for "no impact" on the mean of a specific outcome is a test of no impact for all beneficiaries nationwide.

The impact estimates are "intent to treat" estimates. In other words, they capture the mean impact of the applicability of the benefit offset rules to the earnings of all T1 subjects, whether or not those subjects work and use the offset. Hence, they reflect "no impacts" for the large majority who would not have any

earnings under current law or the offset, as well as for those who fail to learn about, understand, or trust the offset.

Though the method used to produce point estimates of the *impacts* of the benefit offset in this report is exactly the same as that used in the previous Stage 1 *Snapshot Reports*, we have revised our calculations of the standard errors for these estimates. We took this measure to address an issue when calculating standard errors for Stage 2 BOND impacts (see Gubits et al. 2013). The new method addresses the instability in standard error *estimates* that arises from having only 10 BOND study sites. Under certain conditions, the small number of sites can produce standard error estimates that appear to be too small when juxtaposed against other standard error estimates expected to be of similar magnitude. The new method replaces such estimates with an alternative, larger estimate.[9] Full details of all estimation methods appear in Appendix A. The new method does not substantively change the interpretation of findings from previous *Snapshot Reports*.

We make a multiple-comparison adjustment for the two confirmatory outcomes—outcomes selected on the basis of theory and policy interest alone (see Bell et al. 2011). This adjustment is necessary because we are testing multiple outcomes, which makes the probability of a Type I error (rejecting the null hypotheses if it is true) larger than the significance level for the individual tests. To compensate for this effect, we adjust the test statistics for each of the two confirmatory outcomes so that the probability of rejecting the null hypothesis of no impact on either confirmatory outcome is equal to the specified significance level if the null hypothesis is true.[10]

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

We estimate impacts for the overall Stage 1 sample and two pairs of subgroups identified in our *Analysis Plan,* one defined by duration of SSDI benefit receipt and the other by SSI status in the month prior to random assignment (Bell et al. 2011). 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 as three years or less) will 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 use the benefit offset, because they will have completed fewer TWP and GP months at the

---

[9] The new method uses the larger of two standard error estimates for hypothesis testing: (1) a standard error estimate from a model where sites are treated as clusters (errors are correlated within site and independent across sites) and (2) a standard error estimate from a model that includes site fixed effects but does not allow for correlation of errors within site. Standard theoretical statistical analysis implies that the first standard error should be at least as large as the second standard error. Our simulations show that when the true cross-site variance of impacts is relatively small, the first standard error estimate can occasionally be smaller than the second. When the first standard error is observed to be smaller than the second, the new method uses the second estimate for hypothesis testing.

[10] Our approach adjusts the *p*-values for the confirmatory outcomes using the Westfall and Young (1993) method. Details of the *p*-value adjustments for tests of impacts on the confirmatory outcomes appear in Appendix A of this report. See Schochet (2008) for further discussion of the multiple comparisons problem.

outset of the demonstration in comparison to long-duration subjects. The SSI status subgroups are of interest because concurrent subjects face incentives different from those of SSDI-only subjects. For concurrent subjects, SSI benefits are offset more immediately under existing SSI work incentives, and each additional dollar of SSDI benefits is offset by a dollar reduction in SSI benefits after a small disregard. We treat as exploratory all subgroup analyses, including the tests of earnings and SSDI benefits paid.

2.1.4. Final Analysis Sample Sizes

Exhibit 2-2 presents the sizes for the overall sample and the subgroups, none of which have differed across *Snapshot Reports*. The final Stage 1 analysis sample contains a total of 968,713 subjects, spread across T1 (77,115) and C1 (891,598).[11] By design, the T1 sample is evenly split between short- and long-duration beneficiaries (see Bell et al. 2011).[12] Because most SSDI beneficiaries do not receive SSI and sampling rates did not depend on SSI status, there are substantially more SSDI-only subjects relative to concurrent subjects. As a result, we have diminished power to detect impacts for concurrent subjects relative to the other subgroups.

The baseline characteristics (not shown) for the weighted T1 sample are statistically equivalent to those for the weighted C1 sample (Stapleton et al. 2013). These descriptive findings increase our confidence that any statistically significant differences in subsequent outcomes between the T1 and C1groups are "internally valid" impact estimates. That is, significant estimates represent real impacts of the benefit offset on outcomes for the T1 group, rather than systematic pre-existing differences between the two groups or their environments.

---

[11] The team randomly selected nearly 80,000 beneficiaries for the T1 group. The C1 group included an initial core group (C1-core) that was the same size as the T1 group and a supplemental sample of C1 subjects (C1-supplements) that was added to the C1 sample after completion of Stage 2 recruitment (C1-supplement): BOND-eligible subjects who were not included in the samples that were released for Stage 2 recruitment. The C1 sample is the combination of the C1-core and C1-supplement samples. As Stapleton et al. (2013) show, after the application of weights, there were no statistically significant differences between the characteristics of the full C1 sample and those of the C1-core sample. Further, use of the C1-core sample for estimation of impacts on 2011 outcomes yielded estimates that were not statistically different from those based on the full C1 sample. As expected, standard errors were substantially higher when we used only the C1-core sample. Because of the earlier findings for 2011, this report presents findings only from the full C1 group.

[12] As noted in the previous footnote, the C1 group includes C1-Core and C1-Supplemental subjects. The C1-core group was also evenly split between the short- and long-duration groups. The characteristics of the C1 supplemental sample differ because it includes a larger share of long-duration beneficiaries and excludes SSDI-only beneficiaries in the BOND sites that were in the recruitment pool for Stage 2 of BOND. To account for the differing compositions of the T1 and C1 subjects, we use sample weights that eliminate the effects of compositional differences on comparisons of T1 and C1 means. See Stapleton et al. (2013) for more details.

Exhibit 2-2. Stage 1 Analysis Sample Composition

Random Assignment Group	Full Sample	Duration		SSI Status	
		Short Duration	Long Duration	SSDI-Only	Concurrent
T1	77,115	38,669	38,446	64,709	12,406
C1	891,598	209,790	681,808	694,270	197,328

Source: BOND Operations Data System (BODS).

Notes: The Stage 1 analysis sample excludes (1) subjects who were initially assigned to the sample but were later determined to have died prior to assignment, and (2) any pair of beneficiaries (for example, a primary and a DAC or two DACs with the same primary) on a common primary record who were assigned to different BOND groups (see Stapleton et al. 2013 for details on this adjustment). Weights are used to ensure that the BOND subjects who meet the analysis criteria in both the T1 and C1 analysis samples are representative of the national beneficiary population in the month of random assignment. The weighted population size is 6,526,888.

2.2. Previous Findings

The most important finding in previous *Snapshot Reports* (Stapleton et al. 2013, 2014) was the limited evidence of impact on beneficiary earnings. Without positive impacts on earnings, especially on the percentage of beneficiaries with earnings above BYA, the benefit offset will unambiguously increase total SSDI benefits paid by allowing T1 subjects who would earn more than BYA under current program rules to keep a portion of their benefits that they would otherwise lose under current rules.

The previous *Snapshot Reports* also contain evidence of impacts for benefits paid in 2011 and 2012. For 2011, there was a positive, marginally significant impact on SSDI benefits paid. The impact estimate is very small, however—equivalent to $23 per month, representing about 0.3 percent of the total SSDI benefit paid to C1 subjects. For 2012, there was a strongly significant impact on SSDI benefits paid. The dollar magnitude of impact on SSDI benefits paid in 2012 is three times that in 2011, $69 versus $23. This larger value might be attributed at least partially to the greater number of months with the offset in place in 2012: 12 months versus 8 in 2011. For 2012, there was also a 0.05 month increase in the number of months with SSDI payments.

However, earlier reports also showed some evidence of impacts on subgroups of beneficiaries related to earnings and SSI benefit outcomes. In 2011 and 2012, SSDI-only beneficiaries had a lower mean SSI benefit. In 2012, SSDI-only beneficiaries were also more likely to be employed as a result of the offset, while short-duration beneficiaries were less likely to earn above 2 × BYA. In the following sections, we consider whether these significant findings persisted into 2013.

As described in greater detail by Derr et al. (2015), several activities related to BOND's implementation from 2012 through 2014 increased offset usage among T1 subjects that might affect the impact estimates in 2013 reported here. In 2012, the BOND Implementation Team conducted a follow-up outreach effort primarily targeting those with evidence of earnings in 2011. Additional outreach, starting in 2013 and ending in 2014, targeted all T1 subjects who had not yet been in contact with the demonstration.[13] Finally, reconciliation completed in February 2013 using IRS earnings data for 2011, retroactively reduced the 2011 benefits of a substantial number of T1 subjects under the offset. The currently reported results, as

[13] See Derr et al. (2015) for a description of the outreach efforts and their results.

well as effect estimates that appear in later reports, may reflect the consequences of these actions for demonstration impacts.

3. Impact Findings

This chapter presents impact estimates for the confirmatory and exploratory outcomes. As outlined in Chapter 2, each estimate is the difference between the weighted T1 group mean for the outcome and the weighted C1 group mean for the same outcome after statistical adjustments to the C1 group mean for differences in observed characteristics. The T1 sample itself is weighted to reflect the national beneficiary population meeting the BOND eligibility criteria in 2011, so the impact estimates are unbiased estimates of the impact of benefit offset, as implemented under BOND, for the entire eligible population.

Statistically significant impact estimates are identified at the 1, 5, and 10 percent levels. We describe estimates that are statistically significant at the 1 percent level as "strong evidence," at the 5 percent level as "evidence," and at the 10 percent level as "weak evidence." Impact estimates not significant at the 10 percent level or below are considered insignificant. The size of our sample allows our regression-adjusted models to detect very small impacts for several outcomes. As will be seen, some significant impacts are very small as a percentage of the adjusted C1 group mean.

3.1. Full Stage 1 Treatment Group

Exhibit 3-1 presents the estimates of 2013 impacts on earnings and benefit outcomes for the full Stage 1 BOND sample. Total earnings and total benefits paid are the confirmatory outcomes for January through December 2013. All remaining earnings and benefit outcomes are exploratory.

3.1.1. Confirmatory Impacts

We find no evidence that the benefit offset had an impact on earnings in 2013. The mean annual earnings of both T1 and C1 subjects were relatively low ($1,269 and $1,267 for T1 and C1 subjects, respectively), as few subjects in either group (about 13.3 percent) worked during the year. The 95 percent confidence interval for the estimate of $2 ranges from -$43 (negative impact) to $47 (positive impact).

There is strong evidence of a positive impact on SSDI benefits *paid in* 2013. The estimated impact on mean benefits paid in 2013—that is, benefits paid during the 2013 calendar year before any retroactive administrative adjustments were made—was $139, equivalent to 1.2 percent of adjusted mean SSDI benefits paid to C1 subjects during 2013 ($139/$11,232). This 2013 estimate of $139 is twice as large as the estimated impact on mean benefits paid in 2012 ($69). The 95 percent confidence interval for this estimate ranges from $94 to $184.

3.1.2. Exploratory Impacts

For all treatment subjects, the only evidence of an impact for an exploratory benefit outcome is for the number of months with SSDI payments. The mean number of months with SSDI payments is 0.09 higher, equivalent to 0.8 percent of the adjusted mean for C1 subjects (10.69 of a possible 12 months). This finding presumably reflects retention of partial benefits under the offset in some months by T1 subjects who, under current law, would have lost their benefits entirely because of earnings.[14]

[14] For T1 and C1 subjects, the mean total SSI payment was around $380, and the mean number of months with an SSI payment was about 1.7 months. The mean of total SSI benefits paid was small in comparison to the mean of total SSDI benefits paid ($11,512 for C1 subjects), reflecting that only a small minority of Stage 1 subjects received SSI benefits in 2013.

There are no statistically significant impacts for the four exploratory earnings outcomes in Exhibit 3-1. Our findings indicate that most T1 subjects did not work in 2013, and only a small portion had earnings above BYA. About 13 percent of subjects in the T1 and C1 groups have at least some earnings in 2013, including 2.6 percent with earnings above BYA, 1.1 percent with earnings above twice BYA, and 0.6 percent with earnings above three times BYA.

Exhibit 3-1. Stage 1 Impact Estimates on Earnings and Benefit Outcomes

	T1 Mean	C1 Mean	Impact Estimate
Earnings (January–December 2013)			
Total earnings in 2013 **(confirmatory)**	$1,269	$1,267	$2 ($23)
Employment in 2013	13.44%	13.25%	0.19 (0.13)
Earnings above BYA	2.63%	2.54%	0.09 (0.06)
Earnings above 2 × BYA	1.10%	1.17%	-0.07 (0.04)
Earnings above 3 × BYA	0.61%	0.64%	-0.04 (0.03)
Benefits Paid (January–December 2013)			
Total SSDI benefits paid in 2013 **(confirmatory)**	$11,371	$11,232	$139*** ($23)
Number of months with SSDI payments	10.78	10.69	0.09*** (0.02)
Total SSI benefits paid in 2013	$440	$445	-$5 ($8)
Number of months with SSI payments	1.91	1.91	-0.01 (0.01)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria in both the T1 and the C1 analysis samples are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Unweighted sample sizes: T1 = 77,115; C1 = 891,598. See Chapter 2 for variable definitions. Impact estimates are regression-adjusted for baseline characteristics. All outcomes are measured during the second calendar year of operations (January 2013–December 2013). 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 Appendix A for 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 is significantly different from zero at the 10, 5, and 1 percent levels, respectively, using a two-tailed t-test.

3.2. Subgroups

Below we present the impact estimates for the subgroups defined by duration of SSDI benefit receipt (Exhibit 3-2) and SSI status (Exhibit 3-3). The outcome measures are the same as those in Exhibit 3-1, but stratified by subgroup. For reasons outlined in Chapter 2, we consider all subgroup estimates as exploratory outcomes. Hence, we do not adjust significance tests for multiple comparisons. For each pair of subgroups, we first describe adjusted outcome means for C1 subjects in the two subgroups; these values reflect population differences for the subgroups under current law.[15] We then describe impacts within each subgroup of the pair and discuss any evidence of differences in impacts across each pair of subgroups.

3.2.1. Duration Since Award

Consistent with expectations from past research (Liu and Stapleton 2011) and findings for 2011 and 2012 (Stapleton et al. 2013, 2014), there are differences in the *levels* of 2013 earnings and benefits paid for C1 subgroups by duration of prior SSDI receipt (Exhibit 3-2). Short-duration C1 subjects—defined as those on benefits for 36 months or less at baseline—have higher mean earnings ($1,477 versus $1,176) and are more likely to be employed (14.5 versus 12.7 percent) than long-duration C1 subjects. In addition, short-duration subjects have higher SSDI benefit payments than long-duration subjects ($12,102 versus $10,861 for C1 subjects).[16] Total SSI payments are also relatively lower for the short-duration group ($374 versus $475 for C1 subjects).

Although there is no evidence of heterogeneity of impacts across the two subgroups, there are several significant impacts within the duration subgroups. Consistent with the findings for 2012, for the short-duration subgroup there is weak evidence of a decrease in the percentage with earnings above 2 × BYA: -0.12 percentage points. Thus, the number of short-duration T1 subjects induced to lower their earnings to less than 2 × BYA exceeds the number induced to increase their earnings above 2 × BYA. For the long-duration subgroup, there is weak evidence of an increase in employment (we previously found no evidence of an employment impact in the two prior years). The point estimate for the employment impact is 0.31 percentage points, equivalent to 2.5 percent of the adjusted percentage of short-duration C1 subjects who are employed (12.61 percent). As in previously reported findings for 2012, within both subgroups there is evidence of impacts on the number of months with SSDI payments: 0.07 and 0.10 months for short- and-long- duration subjects, respectively.

---

[15] We report only differences in subgroup means that provide at least weak evidence of statistical differences (that is, they are significant at the 10 percent level based on a t-test).

[16] This difference likely reflects the way that SSA indexes pre-SSDI earnings when calculating benefit amounts. Specifically, SSA uses an average wage index (AWI) to inflate past earnings prior to calculating the initial benefit amount; after that, SSA adjusts benefits for inflation each year using a Consumer Price Index (CPI). As the AWI usually increases faster than the CPI, mean benefits for new awardees typically increase every year after adjustment for price inflation.

Exhibit 3-2. Stage 1 Impact Estimates for Subgroups Defined by Duration of SSDI Receipt

	Short Duration			Long Duration			
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	Difference in Impact (7)
Earnings (January–December 2013)							
Total earnings in 2013	$1,459	$1,477	$-18 ($37)	$1,187	$1,176	$11 ($31)	-$29 ($49)
Employment in 2013	14.38	14.48	-0.09 (0.20)	13.03	12.72	0.31* (0.16)	-0.40 (0.25)
Earnings above BYA	3.10	3.04	0.05 (0.10)	2.43	2.32	0.10 (0.11)	-0.05 (0.15)
Earnings above 2 × BYA	1.31	1.43	-0.12* (0.06)	1.01	1.05	-0.04 (0.05)	-0.08 (0.08)
Earnings above 3 × BYA	0.76	0.83	-0.06 (0.05)	0.54	0.56	-0.03 (0.04)	-0.03 (0.07)
Benefits Paid (January–December 2013)							
Total SSDI benefits paid in 2013	$12,202	$12,102	$100** ($41)	$11,011	$10,861	$150*** ($30)	-$50 ($51)
Number of months with SSDI payments	10.79	10.72	0.07** (0.03)	10.77	10.68	0.10*** (0.02)	-0.03 (0.03)
Total SSI benefits paid in 2013	$377	$374	$2 ($8)	$467	$475	$-8 ($11)	$10 ($14)
Number of months with SSI payments	1.48	1.49	-0.01 (0.02)	2.09	2.10	-0.01 (0.02)	0.00 (0.03)

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

Notes: Weights are used to ensure that the BOND subjects who meet analysis criteria in both the T1 and the C1 analysis samples are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Unweighted sample sizes: for short duration, T1 = 38,669 and C1 = 209,790; for long duration, T1 = 38,446 and C1 = 681,808. See Chapter 3 for variable definitions. Impact estimates are regression-adjusted. All outcomes are measured during the second calendar year of operations (January 2013–December 2013). Tests for impacts on all outcomes were conducted independently, without multiple-comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the 10, 5, and 1 percent levels, respectively, using a two-tailed t-test.

3.2.2. SSI Benefit Status

Consistent with expectations and our findings from previous *Snapshot Reports*, there are also large differences in levels of 2013 earnings and benefit levels for C1 subgroups by SSI status (Exhibit 3-3). Relative to concurrent subjects, SSDI-only subjects have higher mean earnings ($1,385 versus $734 for those in C1) and higher mean SSDI benefit payments ($12,473 versus $5,667 for those in C1), which reflects the fact that SSDI-only beneficiaries generally have more substantial earnings histories than concurrent beneficiaries. The percentage employed in 2013 also varied between groups (13.6 versus 11.6 for members of C1). Given that concurrent subjects are classified on the basis of SSI payments at the time of random assignment, it is not surprising that concurrent subjects had substantially higher mean SSI payments in 2013 than did SSDI-only subjects ($2,279 versus $38 for C1 subjects in the two groups, respectively). The fact that some subjects in the SSDI-only group received SSI benefits after random assignment presumably reflects sufficiently large declines in assets or income from other sources to satisfy the SSI means test.

We find no evidence of impacts on the earnings outcomes in 2013. The lack of earnings impacts is generally consistent with previously reported findings for 2012; the only difference is that we found some evidence of a positive impact on employment in 2012. Findings for benefit outcomes are also quite similar to those reported for 2012. There is evidence of positive impacts for months with SSDI payments for both subgroups (though the evidence is weak for the concurrent group) but no evidence of heterogeneity of impacts across subgroups. The impact estimates for the number of months with SSDI payments are 0.09 months for concurrent and SSDI-only subjects. The benefit impacts for SSDI-only subjects, who make up 83.9 percent of all T1 subjects, strongly mirror the impacts shown in Exhibit 3-1 for all T1 subjects. There are no differences in either earnings or payment outcomes by SSI status.

Exhibit 3-3. Stage 1 Impact Estimates for Subgroups Defined by Baseline SSI Status

	SSDI-Only			Concurrent			Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings (January–December 2013)							
Total earnings in 2013	$1,385	$1,385	$0 ($29)	$748	$734	$14 ($31)	-$14 ($42)
Employment in 2013	13.79	13.61	0.18 (0.14)	11.84	11.62	0.21 (0.28)	-0.03 (0.32)
Earnings above BYA	2.90	2.78	0.12 (0.09)	1.42	1.47	-0.04 (0.11)	0.16 (0.14)
Earnings above 2 × BYA	1.28	1.35	-0.07 (0.05)	0.30	0.35	-0.05 (0.05)	-0.02 (0.07)
Earnings above 3 × BYA	0.71	0.76	-0.05 (0.04)	0.13	0.12	0.01 (0.06)	-0.06 (0.07)
Benefits Paid (January–December 2013)							
Total SSDI benefits paid in 2013	$12,607	$12,473	$134*** ($26)	$5,802	$5,667	$135** ($56)	-$1 ($61)
Number of months with SSDI payments	10.84	10.75	0.09*** (0.02)	10.50	10.41	0.09* (0.05)	0.00 (0.05)
Total SSI benefits paid in 2013	$36	$38	$-2 ($2)	$2,258	$2,279	$-21 ($42)	$19 ($42)
Number of months with SSI payments	0.16	0.16	0.00 (<0.01)	9.75	9.82	-0.07 (0.06)	0.07 (0.06)

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

Notes: Weights are used to ensure that the BOND subjects who meet analysis criteria in both the T1 and the C1 analysis samples are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Unweighted sample sizes: for SSDI-only, T1 = 64,709 and C1 = 694,270; for concurrent, T1 = 12,406 and C1 = 197,328. See Chapter 2 for variable definitions. Impact estimates are regression-adjusted. All outcomes are measured during the second calendar year of operations (January 2013–December 2013). Tests for impacts on all outcomes were conducted independently, without multiple-comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the 10, 5, and 1 percent levels, respectively, using a two-tailed t-test.

4. Discussion

In this chapter, we summarize and analyze the impact findings for the confirmatory and exploratory outcomes and compare them with the 2011 and 2012 findings in the previous *Snapshot Reports* (Stapleton et al. 2013, 2014).

4.1. Summary of Findings

The pattern of findings in 2013 for the two confirmatory outcomes (fourth column of Exhibit 4-1) is similar to the pattern in 2011 and 2012: we find no impact on earnings, but continued strong evidence of a positive impact on SSDI benefits *paid in* 2013. The dollar magnitude of impact on SSDI benefits paid in 2013 is almost twice as large as that in 2012. On a per month basis, the impacts on benefits paid in each year are $3 in 2011, $5 in 2012, and $11 in 2013.

Exhibit 4-1. Summary of Impact Findings

	Predicted Sign	Year 1 (May–Dec 2011)	Year 2 (Jan–Dec 2012)	Year 3 (Jan–Dec 2013)
Confirmatory Outcomes				
Total earnings in year	?			
Total SSDI benefits paid (May–Dec only in 2011)	?	$23	$69	$135
Exploratory Outcomes				
Employment and Earnings in Year				
Employment	+			
Earnings above BYA	+			
Earnings above 2 × BYA	?			
Earnings above 3 × BYA	?			
Benefits Paid in Year (May–Dec only in 2011)				
Number of months with SSDI payments	+		0.05	0.09
Total SSI benefits paid	–			
Number of months with SSI payments	–			

Note: The predicted signs are based on economic theory, which abstracts from administrative factors that might have an impact on outcomes (Bell et al. 2011). The estimates from Year 1 are reported in Stapleton et al. (2013). The estimates from Year 2 are reported in Stapleton et al. (2014). The estimates for Year 3 are reported in Exhibit 3-1.

For the exploratory earnings and benefit outcomes for the overall sample, we find no evidence of impacts on any of the employment and earnings outcomes and, as in 2012, strong evidence of an increase in the number of months with SSDI payments. The increase in number of months with SSDI payments (0.09 months) is consistent with the positive impacts shown above for mean SSDI benefits paid in 2013.

We find strong evidence of impacts on the number of months with SSDI payments across all subgroups, but limited, weak evidence of other subgroup impacts and no differences in impacts across subgroup pairs. For short-duration subjects, we find weak evidence of a negative impact on the percentage of subjects with 2013 earnings above 2 × BYA, mirroring a previously reported finding for 2012. For long-duration subjects, we find weak evidence of a positive impact on employment in 2013, but not in previous years.

4.2. Analysis

The confirmatory outcome impact estimates in this third *Snapshot Report* are consistent with those in the previous two snapshot reports—no impact on earnings and an increase in benefits paid as a result of the demonstration. As noted in previous chapters, the earnings finding is critical, because without a positive impact on earnings—especially on the percentage of beneficiaries with earnings above BYA—the benefit offset will lead to higher benefits paid for a given calendar year, relative to benefits paid without the offset. This factor exists because for a given level of earnings, the offset implies a benefit that is at least as high as the current law benefit. Lower benefits can result only if there is a strong increase in earnings, relative to what earnings would have been without the offset. We find no evidence of an increase in earnings. Hence, although the impacts on benefits paid *for* 2013 are likely to differ from the presented impacts on benefits paid *in* 2013, it seems very unlikely that they will be negative. Unlike the benefits paid measure, the earnings measure is not affected by administrative benefit adjustments; therefore, the earnings impacts reported in this and the preceding *Snapshot Reports* will not change.

Although unrealized administrative adjustments to benefits paid will likely make the benefits paid *in* 2013 different from the more policy relevant benefits paid *for* 2013, the finding of an increase in benefits paid *in* 2013 is nevertheless noteworthy for two reasons. First, the 2013 estimate is almost twice as large as the 2012 estimate. Second, potential lack of understanding of the benefit offset among T1 subjects in 2011 and 2012, which might have helped explain the lack of earnings impacts in those years, was presumably at least partially addressed by retroactive benefit adjustments and outreach to T1 subjects in 2013.

The lack of earnings impacts in the first three years might foreshadow no earnings impacts in 2014 or beyond, though three factors related to offset adjustments might affect future earnings responses by T1 subjects. First, there was a 60 percent increase in the number of subjects who received their first offset adjustment in 2014 (after the period for the results reported here) relative to 2013 (from 1.3. to 2.1 percent).[17] As their benefits are adjusted, beneficiaries' understanding of the offset and how to take advantage of it may improve. Second, theory and prior evidence suggest that some C1 subjects will reduce their earnings to less than the cash cliff amount after they discover that their benefits are suspended, and this phenomenon may have been delayed because of work Continuing Disability Review backlogs.[18] Finally, early findings from Stage 2 BOND evaluation (Gubits et al. 2014) indicate positive earnings impacts for that smaller and self-selected subset of SSDI beneficiaries—the 5.4 percent of SSDI-only beneficiaries who volunteered for Stage 2. It follows that the offset is capable of producing a positive impact on the earnings of at least some T1 subjects.

Impacts on benefits paid *for* 2011, 2012, and 2013 will not be known until SSA completes retroactive adjustments for those years and recovers improper payments. As that process unfolds, benefits paid to T1 subjects *for* 2011, 2012, and 2013 will diminish. At the same time, SSA will be making retroactive

---

[17] This finding is based on all adjustments completed through January 2015. The increase in the percentage with an adjustment after all retroactive adjustments are counted might be smaller; after counting retroactive adjustments made in 2014 and January 2015, the percentage with an adjustment for benefits in 2013 or earlier is 1.8 percent, 0.5 percentage points higher than the 1.3 percent with adjustment actually made by the end of 2013. Additional retroactive adjustments will be made in the future.

[18] Weathers and Bailey (2014) report evidence of earnings reductions following the end of the TWP and grace period for subjects in the Accelerated Benefits demonstration, who were all subject to current-law earnings rules.

adjustments to benefits of some C1 subjects and recovering improper payments, under current law processes. Hence, the difference between mean benefits paid to T1 and C1 subjects *for* these three years will not necessarily shrink. The likely effect of administrative factors on benefits paid *in* each year underlines the importance of the final evaluation's planned focus on impacts on benefits paid *for* the evaluation period, rather than benefits paid *in* the evaluation period.

Future evaluation reports will also present estimates of impacts on an array of other outcomes, including those observed via the 36-month follow-up survey of 10,000 T1 and C1 subjects. Findings from the survey, as well as additional process study findings on the demonstration's implementation and further analysis of offset participation, will aid in the interpretation of the impact findings.

References

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

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

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

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

Hanson, R. H. *The Current Population Survey—Design and Methodology*, Technical Paper 40, Washington, DC: U.S. Bureau of the Census, 1978.Livermore, Gina, Allison Roche, and Sarah Prenovitz. *Work Activity and Use of Employment Supports Under the Original Ticket to Work Regulations: SSI and DI Beneficiaries with Work-Related Goals and Expectations.* Washington, DC: Mathematica Policy Research, 2009.

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. "Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations." *NCEE 2008-4018*. Princeton, NJ: Mathematica Policy Research, 2008.

Social Security Administration. *Annual Statistical Report on the Social Security Disability Insurance Program, 2012*. Baltimore, MD: SSA, 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.* Washington, DC: Mathematica Policy Research, 2014.

Stapleton, David, David Wittenburg, Daniel Gubits, David Judkins, David R. 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 C., Stephen H. Bell, David C. Wittenburg, Brian Sokol, and Debi McInnis. *BOND Implementation and Evaluation: BOND Final Design Report.* Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research, December 2010.

Weathers, Robert R. II, and Michelle Stegman Bailey. "The Impact of Rehabilitation and Counseling Services on the Labor Market Activity of Social Security Disability (SSDI) Beneficiaries." *Journal of Public Policy Analysis and Management*, vol. 33, no. 3, 2014, pp. 623–648.

Westfall, Peter H., and S. S. Young. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.* New York: Wiley-Interscience, 1993.

Westfall, Peter H., Randall Tobias, and Russell D. Wolfinger. *Multiple Comparisons and Multiple Tests using SAS.* SAS Institute, 2011.

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, and Washington, DC: Mathematica Policy Research, May 2012.

Wolter, K. M. *Introduction to Variance Estimation*. New York: Springer-Verlag, 1985.

Wright, Debra, Gina Livermore, Denise Hoffman, Eric Grau, and Maura Bardos. *2010 National Beneficiary Survey: Methodology and Descriptive Statistics.* Washington, DC: Mathematica Policy Research, 2011.

Appendix A: Detailed Summary of Methodological Approach

This appendix describes the methods used to estimate impacts. Specifically, we describe the estimation procedure, the multiple comparisons adjustment, and the covariates included in the estimation model. The estimation of standard errors presented in this report differs from earlier Stage 1 *Snapshot Reports*. The new method of estimating standard errors also requires a refinement in the implementation of the multiple comparisons procedure. We note these changes in Section A.1.

A.1. Changes from Previous Stage 1 *Snapshot Reports*

The method used to estimate the *impacts* of the benefit offset in this report is the same as that used in the previous Stage 1 *Snapshot Reports*. However, this report uses a new method to estimate the *standard errors* of the impacts.

The evaluation team developed this new method to address an issue that first arose during impact estimation for Stage 2. During the preparation of the *First- and Second-Year Snapshot Report for Stage 2*, we observed widely varying standard error estimates for pairs of estimates for which we expected the standard errors to be similar. These estimates were calculated using a model similar to that used in the first two Stage 1 *Snapshot Reports*, one in which errors are correlated within site and independent across sites. Upon investigation of the issue, the evaluation team determined that the instability in standard error estimates was related to having a small number of sites—only 10. Our simulations showed that an analysis with 10 sites was vulnerable to unstable variance estimates when true cross-site variance in impacts is relatively small. To address this issue, the evaluation team developed a method in the spirit of Hanson (1978) and Wolter (1985) in which techniques were used to "stabilize" estimates of variance. Because Stage 1 was theoretically as vulnerable to unstable standard error estimates as Stage 2, the evaluation team decided to adopt this new method for both Stage 1 and Stage 2.

The new method uses the larger of two standard error estimates for hypothesis testing: (1) a standard error estimate from a model where errors are correlated within site and independent across sites (that is, an "unconditional" standard error that treats the sites as clusters, sometimes called the "cluster-robust" standard error) and (2) a standard error estimate from a model that includes site fixed effects but does not allow for correlation of errors within site (that is, a "conditional" standard error, sometimes called the "robust, unclustered" standard error.) These two standard errors are appropriate for different circumstances. The unconditional standard errors are designed to support inferences about what would happen with a national implementation of the benefit offset. In contrast, the conditional standard errors are designed to support inferences about what would happen if the benefit offset were implemented throughout the 10 sites. Standard theoretical statistical analysis implies that the true unconditional standard errors are at least as large and usually (often considerably) larger than the conditional standard errors. This situation exists because unconditional inference requires us to extrapolate from the 10 sites to the rest of the nation. However, the estimated (not true) unconditional standard errors are noisy (unstable),

because they use observed variation among a relatively small number of sites. In the new method, we stabilize the unconditional standard errors by replacing them with corresponding conditional standard errors whenever the unconditional standard error is smaller than the conditional standard error.[19]

The new method for estimating standard errors necessitates a corresponding refinement in the multiple comparisons procedure used to adjust the p-values of the two confirmatory significance tests. As in the previous Stage 1 *Snapshot Reports*, we use the Westfall-Young permutation step-down method as the multiple comparisons procedure in this report. This approach involves generating a large number of re-randomized samples and comparing the two p-values (for the two confirmatory outcomes of annual earnings and annual SSDI benefits) in each re-randomized sample with the two p-values from the actual Stage 1 sample. Because the p-values from the actual Stage 1 sample are now derived from the larger of either unconditional or conditional standard errors, the p-values from the re-randomized samples need to be defined such that they use the same standard error as chosen in the actual Stage 1 sample. For example, suppose that in the actual Stage 1 sample, the unconditional standard error for the annual earnings impact is larger than the conditional standard error, while the conditional standard error is larger for the annual SSDI benefits impact. Then, in each re-randomized sample, the p-value for the earnings impact needs to be derived from the unconditional standard error and the p-value for the benefits impact needs to be derived from the conditional standard error.[20]

Below we provide details on the econometric model that is the basis for all impact estimates in the Stage 1 BOND evaluation. Specifically, we describe the estimation procedure, the multiple comparisons procedure, and covariates included in the estimation model.

A.2. Estimation Procedure

We begin our description of the approach with the general estimation model in Equation (1) and then follow with the detailed specification used in this report in Equation (3). The general estimation model under this approach is:

(1) $$y_{ij} - \hat{y}_{ij} = \beta_0 + \beta_1 T_{1ij} + \varepsilon_{ij},$$

where y_{ij} is an outcome measure for beneficiary *i* in site *j* (*j* = 1, 2, …, 10),
$\hat{y}_{ij}$ = the predicted outcome for beneficiary *i* in site *j*,
T_{1ij} = an indicator of whether beneficiary *i* in site *j* has been randomized into the T1 group (= 1 if so, = 0 if in C1 group),
β_0 = the model intercept,
β_1= the overall impact of the T1 treatment (versus the no treatment of the C1 group), and
ε_{ij} is an error term that is correlated within site and independent between sites:

---

[19] Our simulations have shown that the likelihood of the conditional standard error being larger than the unconditional standard error increases as the true cross-site variance of impacts decreases. In a simulation of very small true cross-site variance of impacts, we found a Type I error rate of only 0.017 when using $\alpha = 0.05$. This result shows that when true cross-site variance is relatively small (and so occasionally the conditional standard error is larger than the unconditional standard error), the variance stabilization method is conservative, sacrificing some statistical power to avoid displaying grossly inconsistent variance estimates for pairs of statistics where generally similar variances are expected.

[20] Our simulations show that this refinement controls the family-wise error rate at the expected level.

$$cov\left(\varepsilon_{ij}, \varepsilon_{i'j'}\right) = \begin{cases} \varphi^2 | i = i', j = j' \\ \rho\varphi^2 | i \neq i', j = j'. \\ 0 | j \neq j' \end{cases}$$

The predicted outcome $\hat{y}_{ij}$ is calculated from a first-stage regression model (a "working model"):

(2) $y_{ij} = \alpha_0 + X_{ij}\tilde{\alpha}_1 + \mu_{ij}$,

where y_{ij} is defined as above,
X_{ij} = a vector of baseline characteristics for individual *i* in site *j*,
α_0 = the model intercept,
$\tilde{\alpha}_1$ = a vector of coefficients, and
μ_{ij} is an i.i.d. distributed error term.

This first-stage regression is estimated on the C1 group only. The parameter estimates are then used to calculate the predicted outcome ($\hat{y}_{ij}$) for both T1 and C1 beneficiaries. Subtracting the predicted outcome from the actual outcome serves to remove the variation in the outcome that can be explained by the covariates. The residuals that are produced may then be analyzed to measure the impact of BOND (that is, being assigned to T1 rather than to C1), as in Equation (1).

Rather than directly analyzing the residuals, however, we add a step to reduce the size of the data. This data reduction accomplishes two purposes: (1) it greatly speeds the run time of the multiple comparisons adjustment and (2) it appropriately addresses the non-normal distributions of earnings and binary outcomes. To achieve this data reduction, we split each "site X assignment group" cell into 200 evenly sized random groups. For instance, the T1 group in the Alabama site is randomly split into 200 groups, and the C1 group in Alabama is also randomly split into 200 groups. This approach results in 4,000 random groups (10 sites × 2 assignment groups × 200 random groups). Within each random group, the average residual is computed, and the group's weight is the sum of the weights of its members. These average residuals are then used to calculate the impact estimate.[21]

This data reduction speeds our multiple comparisons procedure, which is based on resampling, because repeated computer processing of 4,000 observations is faster than repeated processing of roughly 970,000 observations. The data reduction also serves to address the non-normal distributions of the earnings outcome and binary outcomes. Given the non-normality of these outcomes, the residuals of individual beneficiaries violate normality. However, the central limit theorem ensures that the distribution of *average* residual is normal, even if the individual residuals are not normally distributed. This fact makes the data-reduction step appealing on statistical grounds.

Incorporating the data reduction into our approach results in the two following slightly different estimation models used in this report:

[21] This average residual is calculated using sampling weights, so that beneficiaries with higher sampling weights make a larger contribution to the average residual.

(3) $\bar{R}_{kaj} = \beta_0 + \beta_1 T_{1kaj} + \varepsilon_{kaj}$,

(4) $\bar{R}_{kaj} = \gamma_0 + \gamma_1 T_{1kaj} + \omega_j + \tau_{kaj}$,

where $\bar{R}_{kaj} = \frac{1}{\sum_{m=1}^{n_{kaj}} w_m} \sum_{m=1}^{n_{kaj}} w_m (y_m - \hat{y}_m)$, the weighted average residual over the n_{kaj} members of random group k within assignment group a (either T1 or C1) in site j,

w_m= the sampling weight of beneficiary m of the random group indexed by kaj,

T_{1kaj} = an indicator of whether the members of random group k within assignment group a in site j have been randomized into the T1 group (= 1 if so, = 0 if in C1 group),

β_0, γ_0 = weighted averages of site-specific intercepts,

β_1, γ_1 = weighted averages of site-specific impacts of the T1 treatment (versus the no treatment of the C1 group),

ε_{kaj} is an error term that is correlated within site and independent between sites:

$$cov\left(\varepsilon_{kaj}, \varepsilon_{k'a'j'}\right) = \begin{cases} \varphi^2 | k = k', a = a', j = j' \\ \rho\varphi^2 | k \neq k' or\ a \neq a', j = j', \\ 0 | j \neq j' \end{cases}$$

ω_j is a site-specific fixed effect for site j, and

τ_{kaj} is an error term that is independent within site and between sites:

$$cov\left(\tau_{kaj}, \tau_{k'a'j'}\right) = \begin{cases} \varphi^2 | k = k', a = a', j = j' \\ 0 | k \neq k' or\ a \neq a'\ or\ j \neq j' \end{cases}.$$

The estimation of Equation (3) produces the estimates of the impact (β_1) and the unconditional standard error. The estimation of Equation (4) produces the estimate of the conditional standard error. The larger of the unconditional and conditional standard errors is used for hypothesis testing (and is presented in report exhibits). The estimation of Equations (3) and (4) both incorporate the weights of the random groups, to produce nationally representative results. We estimate both equations using the PROC SURVEYREG procedure in the SAS software package.[22] Equation (3) is estimated using the site variable in the CLUSTER statement. Equation (4) is estimated using the site variable in the STRATUM statement and including site dummy variables.

[22] We note that the estimated standard errors for the intervention impact produced by the PROC SURVEYREG procedure do not take into account uncertainty in the estimates of the $\tilde{\alpha}_1$ parameters in Equation (2). This factor has the potential to bias the estimates of standard errors downward, but we estimated the bias was very small (less than 1 percent), primarily because of the large sample sizes in BOND. Prior to running the final specifications at SSA, we estimated the standard error for the impact on SSDI benefits using an alternative jackknife estimator that captured the uncertainty in the estimates of the $\tilde{\alpha}_1$ parameters in Equation (2). We found the downward bias was too small to measure. For example, in one of our benefit equations, we estimated that the jackknife procedure reduced the standard error by $0.03, which was less than one percent of the standard error without the correction. This evidence, in addition to the additional run time that would result from the use of the jackknife estimator in conjunction with our multiple comparisons procedure, led us to the decision not to use the jackknife estimator for impact estimation for all estimates.

Previous Stage 1 Snapshot Reports uniformly used the estimated standard errors for Equation (3), rather than the larger standard error from Equation (3) or Equation (4). Exhibit A-1 shows that relative to the original estimation procedure, the new estimation procedure had no effect on the 2013 confirmatory outcome impact estimates, associated standard errors, and unadjusted p-values. For other 2013 outcomes, the standard errors may vary between methods, with the new method always calculating a standard error at least as large as the old method. The very small changes in adjusted p-values are probably due to chance variation in the permutation streams.

Exhibit A-1. Stage 1 Impact Estimates on 2013 Confirmatory Outcomes Illustrating Change in Standard Error Estimation and Refinement of Multiple Comparisons Adjustment

	Impact Estimate: T1 vs. C1 (1)	Standard Error (2)	*p*-value (Unadjusted) (3)	*p*-value (Multiple Comparisons Adjustment) (4)
Methods Used in Previous Stage 1 Snapshot Reports				
Total 2013 earnings **(confirmatory)**	$2	$23	0.919	0.923
Total 2013 SSDI benefits paid **(confirmatory)**	$139	$23	<0.001	0.001
New Methods of Standard Error Estimation and Multiple Comparisons Adjustment				
Total 2013 earnings **(confirmatory)**	$2	$23	0.919	0.918
Total 2013 SSDI benefits paid **(confirmatory)**	$139	$23	<0.001	<0.001

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

Notes: Weights are used to ensure that the BOND subjects who meet analysis criteria in both the T1 and C1 analysis samples are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Unweighted sample sizes: T1 = 77,115; C1 = 891,598. Impact estimates are regression-adjusted for baseline characteristics. 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. The unadjusted *p*-value in the fourth column shows the statistical test prior to the multiple comparison adjustment. The adjusted *p*-value in the fifth column shows the statistical test after the multiple comparison adjustment.

*/**/*** Impact estimate is significantly different from zero at the 0.10/0.05/0.01 levels, respectively, using a two-tailed t-test.

A.3. Multiple Comparisons Procedure

The BOND impact analysis involves running a large number of hypothesis tests due to the inclusion of a large number of outcome measures to be examined and the analysis of numerous subgroups. Having such a large number of hypothesis tests creates a danger of "false positives" arising in the analysis, that is, of finding statistically significant impacts for some outcomes when in fact the true impact of BOND on these outcomes is zero. This danger is called the "multiple comparisons problem." The probability of finding a false positive rises as the number of hypothesis tests performed rises. Given the large number of hypothesis tests to be in BOND, it is very likely that there will be one or more such false positives.

The impact analysis takes two measures to address the multiple comparisons problem in the BOND impact analysis. First, prior to the impact analysis, the hypothesis tests are separated into "confirmatory" and "exploratory" tests, as specified in Bell et al. (2011). Only the two most important outcomes from the evaluation—total earnings and total SSDI benefits paid—are included in the confirmatory group.[23] All other impact estimates, including all estimates for subgroups, are considered exploratory. Statistically significant findings from confirmatory analyses are interpreted as evidence that the benefit offset had impacts on these outcomes, without cause for concern that they reflect the multiple comparisons problem. In contrast, statistically significant findings from exploratory analyses that do not adjust for multiple comparisons are characterized as suggestive of what BOND can accomplish, but might simply reflect the fact that a few impact estimates are bound to be significant when impacts on a large number of outcomes are tested, even if there is no impact on any outcome.

Second, we implement a multiple comparisons adjustment procedure for our two confirmatory outcomes. The procedure accounts for a "family-wise error rate," which represents the probability of rejecting at least one null hypothesis in a family of hypothesis tests when all null hypotheses are true.

For our set of confirmatory tests (tests of the statistical significance of impact estimates for total earnings and total SSDI benefits), the family-wise error rate is defined as the probability of finding a significant impact on either total earnings or total SSDI benefits when the true impact on both outcomes is zero. We employ a method from Westfall and Young (1993) called the permutation stepdown method.[24] In conjunction with the estimation procedure described in A.1, the permutation stepdown method involves reassigning the 4,000 random groups to T1 or C1 many times (20,000) and recalculating impacts on earnings and SSDI benefits each time. In a large-scale simulation of the permutation step-down method using our estimation procedure, we found that this method rejected null hypotheses at the expected family-wise error rate (that is, this method provided the desired protection against false positives).

The permutation step-down method produces adjusted *p*-values for the impacts on total earnings and total SSDI benefits. We describe the method below:

In notation, let

A, B = two outcomes of interest (in this case, earnings and SSDI benefits).

p_A^{raw}, p_B^{raw} = *p*-values from individual t-tests of impact estimates based on the variance-stabilized standard error calculation (the larger of either the standard error assuming an unconditional error term or the conditional standard error). These are the "raw," unadjusted *p*-values for each outcome.

We can then place the outcomes in the order of their raw *p*-values.

[23] The BOND Snapshot reports and interim reports will contain findings for varying lengths of time. In each report, impacts on total earnings and total SSDI benefits for the periods covered will be treated as confirmatory.

[24] This method is also described in Westfall et al. (2011).

IMPACT1, IMPACT2 = the outcomes in order of their raw *p*-values. OUTCOME1 is the outcome with the smaller raw *p*-value, and OUTCOME2 is the outcome with the larger raw *p*-value.

$p^{raw}_{IMPACT1}, p^{raw}_{IMPACT2}$ = 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 two outcomes, producing two *p*-values. (For each impact, the estimation of the replicate sample will employ the same type of standard error—unconditional or conditional—as selected by the variance stabilization procedure.[25])

We can then define the adjusted *p*-values as follows:

$$p^{adj}_{IMPACT1} = \frac{Number\ of\ replicates\ where\ \min\{p^{rep}_{IMPACT1}, p^{rep}_{IMPACT2}\} < p^{raw}_{IMPACT1}}{R},$$

$$p^{adj}_{IMPACT2} = \max\left\{p^{adj}_{IMPACT1}, \frac{Number\ of\ replicates\ where\ p^{rep}_{IMPACT2} < p^{raw}_{IMPACT2}}{R}\right\},$$

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

The *p*-values shown in this report for the confirmatory outcomes of annual earnings and annual SSDI benefits are the adjusted *p*-values calculated using this permutation step-down procedure.

Exhibit A-1 shows that relative to the original estimation procedure, the new estimation procedure had no effect on the 2013 confirmatory outcome impact estimates' unadjusted p-values. The very small changes in adjusted p-values are probably due to chance variation in the permutation streams.

.

[25] This refinement to the multiple comparisons procedure was necessitated by the change in method of standard error estimation.

A.4. Covariates

Exhibit A-2 lists the covariates included in the estimation of Equation (2) in Section A.2. This list of covariates is identical to that used in previous Stage 1 *Snapshot Reports*, and is included here for convenience.

Exhibit A-2. Covariates Included in the Estimation Procedure

Covariates (measured at baseline unless otherwise specified)
Age
Age (squared)
AIME (Average Indexed Monthly Earnings) as of May 2011
AIME as of May 2011 (squared)
AIME as of May 2011 are equal to zero
Any employment in 2010 (the year prior to random assignment year)[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 (the year prior to Random Assignment year)[a]
Gender
Has a representative payee
Has auxiliary beneficiary (AUX) who *is not* a DAC or DWB
Has SSDI start date on or after January 1, 2010 (very short-duration beneficiary)
Ineligible for Stage 2 for geographical reasons
Ineligible for Stage 2 for having a legal guardian who was not a representative payee
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
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
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
Receives written beneficiary notices in Spanish
Rural area dummy
Short-duration SSDI receipt (36 months or fewer)
SSI receipt dummy

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

A.5. Sample Adjustments and Analysis Weights

The adjustments to the Stage 1 sample and the analysis weights used in this report are identical to previous Stage 1 *Snapshot Reports*. Full details about sample and weight construction are provided in the appendix of the first Stage 1 *Snapshot Report* (Stapleton et al. 2013).

Appendix B: 2011 and 2012 Impact Estimates Using Revised Methodology

In this appendix, we apply the updated standard error calculation methodology described in Appendix A to the impact estimates from the 2011 and 2012 *Snapshot Reports*. When compared with the impact estimates in the previous *Snapshot Reports*, the impact estimates in this appendix show that the differences obtained from applying the two different methodologies are minimal.

B.1. 2011 Impact Estimates

Exhibit B-1. 2011 Stage 1 Impact Estimates on Earnings and Benefit Outcomes

	T1 Mean	C1 Mean	Impact Estimate
Earnings (January–December 2011)			
Total earnings in 2011 **(confirmatory)**	$1,195	$1,204	$-9 ($25)
Employment in 2011	16.15	16.03	0.13 (0.12)
Earnings above BYA	2.43	2.41	0.02 (0.12)
Earnings above 2 × BYA	0.95	0.97	-0.03 (0.05)
Earnings above 3 × BYA	0.53	0.53	0.00 (0.03)
Benefits Paid (May–December 2011)			
Total SSDI benefits paid in 2011 **(confirmatory)**	$7,531	$7,508	$23* ($11)
Number of months with SSDI payments	7.49	7.49	0.00 (<0.01)
Total SSI benefits paid in 2011	$340	$342	$-2 ($6)
Number of months with SSI payments	1.37	1.38	0.00 (<0.01)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria in both the T1 and the C1 analysis samples are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Unweighted sample sizes: T1 = 77,115; C1 = 891,598. See Chapter 2 for variable definitions. Impact estimates are regression-adjusted for baseline characteristics. Benefit outcomes are measured for the period from the date of random assignment (May 1, 2011) through December 2011, whereas employment and earnings outcomes are for the full calendar year, including the four months before random assignment. 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 Appendix A for 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 is significantly different from zero at the 10, 5, and 1 percent levels, respectively, using a two-tailed t-test.

Exhibit B-2. 2011 Stage 1 Impact Estimates for Subgroups Defined by Duration of SSDI Receipt

	Short Duration			Long Duration			Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings (January–December 2011)							
Total earnings in 2011	$1,300	$1,337	$-37 ($40)	$1,149	$1,146	$3 ($29)	$-40 ($49)
Employment in 2011	16.80	16.73	0.06 (0.23)	15.88	15.72	0.15 (0.14)	-0.09 (0.27)
Earnings above BYA	2.69	2.75	-0.06 (0.10)	2.32	2.27	0.05 (0.13)	-0.11 (0.17)
Earnings above 2 × BYA	1.11	1.20	-0.09 (0.07)	0.88	0.88	0.00 (0.05)	-0.09 (0.09)
Earnings above 3 × BYA	0.68	0.70	-0.02 (0.08)	0.47	0.46	0.01 (0.03)	-0.03 (0.08)
Benefits Paid (May–December 2011)							
Total SSDI benefits paid in 2011	$8,300	$8,270	$30 ($19)	$7,198	$7,180	$18 ($14)	$12 ($24)
Number of months with SSDI payments	7.57	7.57	0.00 (0.01)	7.46	7.46	0.00 (0.01)	0.00 (0.02)
Total SSI benefits paid in 2011	$368	$376	$-8 ($10)	$328	$327	$1 ($7)	$-9 ($12)
Number of months with SSI payments	1.09	1.09	0.00 (<0.01)	1.50	1.50	-0.01 (<0.01)	0.01 (0.01)

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

Notes: Weights are used to ensure that the BOND subjects who meet analysis criteria in both the T1 and the C1 analysis samples are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Unweighted sample sizes: for short duration, T1 = 38,669 and C1 = 209,790; for long duration, T1 = 38,446 and C1 = 681,808. See Chapter 3 for variable definitions. Impact estimates are regression-adjusted. Benefit outcomes are measured for the period from the date of random assignment (May 1, 2011) through December 2011, whereas employment and earnings outcomes are for the full calendar year, including the four months before random assignment. Tests for impacts on all outcomes were conducted independently, without multiple-comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the 10, 5, and 1 percent levels, respectively, using a two-tailed t-test.

Exhibit B-3. 2011 Stage 1 Impact Estimates for Subgroups Defined by Baseline SSI Status

	SSDI-Only			Concurrent			Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings (January–December 2011)							
Total earnings in 2011	$1,302	$1,308	$-6 ($31)	$713	$735	$-22 ($25)	$16 ($40)
Employment in 2011	16.30	16.13	0.17 (0.13)	15.50	15.57	-0.06 (0.28)	0.23 (0.31)
Earnings above BYA	2.71	2.66	0.06 (0.13)	1.16	1.31	-0.15 (0.10)	0.21 (0.17)
Earnings above 2 × BYA	1.12	1.14	-0.02 (0.07)	0.17	0.22	-0.06 (0.05)	0.04 (0.08)
Earnings above 3 × BYA	0.63	0.63	0.00 (0.04)	0.07	0.06	0.01 (0.03)	-0.01 (0.05)
Benefits Paid (May–December 2011)							
Total SSDI benefits paid in 2011	$8,376	$8,356	$20 ($13)	$3,726	$3,696	$31 ($24)	$-11 ($27)
Number of months with SSDI payments	7.54	7.54	0.00 (<0.01)	7.29	7.26	0.03 (0.03)	-0.03 (0.03)
Total SSI benefits paid in 2011	$33	$37	$-4* ($2)	$1,724	$1,714	$10 ($30)	$-14 ($30)
Number of months with SSI payments	0.07	0.07	0.00 (<0.01)	7.25	7.28	-0.02 (0.03)	0.02 (0.03)

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

Notes: Weights are used to ensure that the BOND subjects who meet analysis criteria in both the T1 and the C1 analysis samples are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Unweighted sample sizes: for SSDI-only, T1 = 64,709 and C1 = 694,270; for concurrent, T1 = 12,406 and C1 = 197,328. See Chapter 2 for variable definitions. Benefit outcomes are measured for the period from the date of random assignment (May 1, 2011) through December 2011, whereas employment and earnings outcomes are for the full calendar year, including the four months before random assignment. Tests for impacts on all outcomes were conducted independently, without multiple-comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the 10, 5, and 1 percent levels, respectively, using a two-tailed t-test.

B.2. 2012 Impact Estimates

Exhibit B-4. 2012 Stage 1 Impact Estimates on Earnings and Benefit Outcomes

	T1 Mean	C1 Mean	Impact Estimate
Earnings (January–December 2012)			
Total earnings in 2012 **(confirmatory)**	$1,283	$1,277	$6 ($31)
Employment in 2012	15.41	15.34	0.06 (0.13)
Earnings above BYA	2.68	2.60	0.08 (0.11)
Earnings above 2 × BYA	1.12	1.12	-0.01 (0.08)
Earnings above 3 × BYA	0.58	0.62	-0.04 (0.03)
Benefits Paid (January–December 2012)			
Total SSDI benefits paid in 2012 **(confirmatory)**	$11,393	$11,324	$69*** ($17)
Number of months with SSDI payments	11.03	10.98	0.05*** (0.01)
Total SSI benefits paid in 2012	$459	$462	$-3 ($8)
Number of months with SSI payments	1.99	1.99	0.00 (0.01)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria in both the T1 and the C1 analysis samples are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Unweighted sample sizes: T1 = 77,115; C1 = 891,598. See Chapter 2 for variable definitions. Impact estimates are regression-adjusted for baseline characteristics. All outcomes are measured during the second calendar year of operations (January 2012–December 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 Appendix A for 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 is significantly different from zero at the 10, 5, and 1 percent levels, respectively, using a two-tailed t-test.

Exhibit B-5. 2012 Stage 1 Impact Estimates for Subgroups Defined by Duration of SSDI Receipt

	Short Duration			Long Duration			Difference in Impact (7)
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	
Earnings (January–December 2012)							
Total earnings in 2012	$1,418	$1,418	$0 ($46)	$1,225	$1,215	$10 ($33)	-$10 ($57)
Employment in 2012	15.60	15.58	0.02 (0.22)	15.32	15.24	0.08 (0.17)	-0.06 (0.28)
Earnings above BYA	3.02	2.95	0.07 (0.14)	2.53	2.44	0.09 (0.12)	-0.02 (0.18)
Earnings above 2 × BYA	1.20	1.35	-0.15* (0.07)	1.08	1.03	0.05 (0.09)	-0.20 (0.12)
Earnings above 3 × BYA	0.75	0.78	-0.03 (0.08)	0.51	0.55	-0.04 (0.03)	0.01 (0.09)
Benefits Paid (January–December 2012)							
Total SSDI benefits paid in 2012	$12,310	$12,269	$41 ($23)	$10,996	$10,919	$77*** ($23)	$-36 ($33)
Number of months with SSDI payments	11.08	11.05	0.03 (0.02)	11.01	10.95	0.06** (0.02)	-0.03 (0.03)
Total SSI benefits paid in 2012	$404	$409	$-5 ($10)	$482	$485	$-3 ($11)	-$2 ($15)
Number of months with SSI payments	1.55	1.56	-0.01 (0.01)	2.17	2.18	-0.01 (0.01)	0.00 (0.02)

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

Notes: Weights are used to ensure that the BOND subjects who meet analysis criteria in both the T1 and the C1 analysis samples are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Unweighted sample sizes: for short-duration, T1 = 38,669 and C1 = 209,790; for long-duration, T1 = 38,446 and C1 = 681,808. See Chapter 3 for variable definitions. Impact estimates are regression-adjusted. All outcomes are measured during the second calendar year of operations (January 2012–December 2012). Tests for impacts on all outcomes were conducted independently, without multiple-comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the 10, 5, and 1 percent levels, respectively, using a two-tailed t-test.

Exhibit B-6. 2012 Stage 1 Impact Estimates for Subgroups Defined by Baseline SSI Status

	SSDI-Only			Concurrent			
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	Difference in Impact (7)
Earnings (January–December 2012)							
Total earnings in 2012	$1,393	$1,382	$11 ($38)	$792	$802	$-10 ($34)	$21 ($51)
Employment in 2012	15.56	15.38	0.18 (0.15)	14.74	15.18	-0.46 (0.36)	0.64 (0.39)
Earnings above BYA	2.92	2.82	0.10 (0.13)	1.57	1.61	-0.04 (0.12)	0.14 (0.17)
Earnings above 2 × BYA	1.31	1.30	0.01 (0.10)	0.26	0.34	-0.08 (0.06)	0.09 (0.12)
Earnings above 3 × BYA	0.68	0.73	-0.05 (0.04)	0.12	0.10	0.02 (0.05)	-0.07 (0.07)
Benefits Paid (January–December 2012)							
Total SSDI benefits paid in 2012	$12,650	$12,588	$62*** ($19)	$5,731	$5,651	$80 ($52)	-18 ($55)
Number of months with SSDI payments	11.09	11.04	0.05** (0.01)	10.76	10.67	0.09* (0.05)	-0.04 (0.05)
Total SSI benefits paid in 2012	$36	$38	$-2 ($3)	$2,364	$2,375	$-11 ($41)	$9 ($41)
Number of months with SSI payments	0.13	0.13	0.00 (<0.01)	10.33	10.34	-0.01 (0.04)	0.01 (0.04)

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

Notes: Weights are used to ensure that the BOND subjects who meet analysis criteria in both the T1 and the C1 analysis samples are representative of the national beneficiary population in the month of random assignment. Standard errors are in parentheses. Unweighted sample sizes: for SSDI-only, T1 = 64,709 and C1 = 694,270; for concurrent, T1 = 12,406 and C1 = 197,328. See Chapter 2 for variable definitions. Impact estimates are regression-adjusted. All outcomes are measured during the second calendar year of operations (January 2012–December 2012). Tests for impacts on all outcomes were conducted independently, without multiple-comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the 10, 5, and 1 percent levels, respectively, using a two-tailed t-test.