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these estimates have been updated over the course of the demonstration, making the published estimates in this report out of date.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

BOND Implementation and Evaluation

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

Deliverable 24c.1

May 28, 2013

Submitted To:
Social Security Administration
Attn: Ms. Joyanne Cobb
Office of Program Development and Research
6401 Security Boulevard
Altmeyer Building, Room 128
Baltimore, Maryland 21235

Contract No. SS00-10-60011

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

Report Context

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

The BOND project includes two stages. Stage 1 is designed to examine how a national benefit offset would affect earnings and program outcomes for the entire SSDI population. Stage 2 is designed to learn more about impacts for those most likely to use the offset (recruited and informed volunteers) and to determine the extent to which significant enhancements to counseling services affect impacts.

This document is the fourth report for the evaluation and the second focused on Stage 1. Two earlier reports provide important reference material about the demonstration design (Stapleton et al. 2010) and the evaluation plan (Bell et al. 2011), including the anticipated outcomes of the demonstration. A third report assessed early implementation activities and provided information on Stage 1 subjects (Wittenburg et al. 2012).

This *Snapshot Report*, which is intended to provide a brief presentation of intermediate results, documents impacts on earnings and benefit outcomes—that is, earnings under the benefit offset relative to earnings under current rules—during the year the demonstration was launched, 2011. The report compares benefit and employment outcomes for all Stage 1 treatment subjects (T1) to those for control subjects (C1). Given the midyear launch of the demonstration and the time necessary for beneficiaries to respond, impacts during the period covered by this report were expected to be small and then grow in subsequent years. The report is the first in a series of annual reports that will track impacts through 2017. The evaluation team will produce a parallel series of *Snapshot Reports* for Stage 2.

Summary of Key Findings

For the eight months of calendar year 2011 after random assignment, we found no evidence that the benefit offset had impacts on the primary outcomes of total earnings and total SSDI benefits paid. Statistically significant but small impacts were found for other outcomes and some subgroups. The lack of substantial impact findings for this period is not surprising given the anticipated trajectory of impacts (Stapleton et al. 2010; Bell et al. 2011). Future evaluation reports will document how benefit offset impacts change annually through 2017.

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. These individuals have no role in implementing or overseeing the BOND intervention they are studying, nor do any members of their evaluation team. Separation of implementation and evaluation does not extend throughout the project, however. 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 rests with the evaluation co-directors. David Stapleton led the writing of this report.

Table of Contents

Acronyms Used in This Report

AIME	Average Indexed Monthly Earnings
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
EWIC	Enhanced Work Incentive Counseling
GP	Grace Period
HLM	Hierarchical Linear Model
IRWE	Impairment Related Work Expense
IRS	Internal Revenue Service
MBR	Master Beneficiary Record
MEF	Master Earnings File
SER	Summary Earnings Record
SEs	Standard Errors
SGA	Substantial Gainful Activity
SSA	Social Security Administration
SSI	Supplemental Security Income
SSDI	Social Security Disability Insurance
SSR	Supplemental Security Record
TTW	Ticket to Work
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 variants of Social Security Disability Insurance (SSDI) program rules governing work and other supports. This is the first in a series of *Snapshot Reports* about the impacts of the demonstration rules on beneficiary outcomes—most notably earnings and benefits paid.[1] This introductory chapter provides a synopsis of the demonstration, describes the purpose of this report, and ends with an outline of the rest of the report.

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 benefits.[2] Specifically, benefits are lost if an SSDI beneficiary's countable monthly earnings exceed the monthly Substantial Gainful Activity (SGA) amount after completing a nine-month Trial Work Period (TWP) and a three-month Grace Period (GP). In 2011, the SGA amount was $1,000 per month for non-blind beneficiaries and $1,640 per month for blind beneficiaries. The complete loss of benefits for earnings in excess of the SGA amount is sometimes called the "cash cliff." The cash cliff gives SSDI beneficiaries an incentive to keep earnings below the SGA level—an incentive that is especially strong for those only able to earn somewhat above the SGA amount.

BOND replaces the cash cliff with a ramp, or a benefit offset that is expected to increase the earnings of those who might otherwise keep their earnings below the SGA amount, and, in so doing, increase their household incomes and reduce their benefits. Specifically, BOND changes the accounting period from monthly to annual and replaces the cash cliff with a benefit offset that gradually reduces benefits when earnings surpass the annual equivalent of the SGA amount. The benefit offset reduces benefits by $1 for every $2 in countable annual earnings in excess of the BOND Yearly Amount (BYA) following the completion of the GP. BYA is equal to 12 times the monthly SGA amount (in 2011, $12,000 for non-blind treatment subjects and $19,680 for blind treatment subjects).

BOND includes two stages. The report focuses on the initial impact of the benefit offset, along with certain changes to ancillary supports, on earnings and benefit outcomes for beneficiaries who were randomly assigned to the Stage 1 treatment group. Their outcomes are compared to those for beneficiaries randomly assigned to the Stage 1 control group, who continued to have their benefits adjusted on the basis of current law and current ancillary supports. The changes to ancillary supports include replacement of counseling services originally available from Work Incentives Planning and Assistance (WIPA) grantees with Work Incentive Counseling (WIC) services, designed to be comparable apart from the fact that they were structured around the benefit offset rules and administrative processes.[3] They also include

---

1 These reports are referred to as "letter reports" in the contract based on the original intent of the report, which was to provide SSA with information on impacts, but we have changed the name to *Snapshot Reports* given that these reports will now be distributed to a broad policy audience.

2 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. See Stapleton et al. (2010) for further details.

3 The WIPA program was suspended in June 2012, but will be reinstated starting in August 2013.

administrative changes implemented by SSA and the BOND implementation team.[4] The latter concern processes for notification of treatment subjects and responding to their inquiries; collection and review of earnings and other information to determine TWP status and adjust earnings; and adjustment of benefits.[5] 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 Enhanced Work Incentive Counseling (EWIC) services relative to WIC services. The evaluation team will document the outcomes of Stages 1 and 2 in a series of parallel reports (see Bell et al. 2011 for more details).[6]

The evaluation team is responsible for all of the estimates that appear in this report. In previous reports, we described the BOND design and the framework for estimating the impacts and summarized early assessment activities on the infrastructure to support Stage 1 service delivery (Stapleton et al. 2010; Bell et al. 2011; and Wittenburg et al. 2012, respectively).[7]

1.2. Purpose

This *Snapshot Report* presents estimates of the impacts of the benefit offset, WIC, and other administrative changes for Stage 1 (hereafter referred to as "benefit offset impacts") during the first eight months of implementation, from May 2011 through December 2011. We applied the evaluation analysis framework specified in Bell et al. (2011) to estimate the impacts that appear in this report.[8] Within that framework, the two most important evaluation outcomes, referred to as "confirmatory outcomes," are total earnings and total SSDI benefits. We use SSA administrative data to estimate impacts on benefits and earnings. Statistically significant findings (e.g., higher earnings and lower benefits impacts) can be interpreted as confirming the effectiveness of the benefit offset.

The report also presents estimates for several exploratory outcomes that are measured in the administrative data (for example, an indicator for earnings above BYA). These exploratory findings provide further information on the impacts of the benefit offset over a broader set of benefit and

---

4 SSA retained its adjudicative role in benefit adjustment and related activities, such as verification of earnings information and distribution of benefit checks.

5 More specifically, the administrative changes include: adoption of an annual rather than monthly accounting period to determine the benefit amount; adoption of federal income tax rules for defining annual earnings; prospective estimation of annual earnings and IRWE, with end-of-the-year benefit reconciliation; a demonstration information system to facilitate and expedite earnings reporting; a centralized, largely automated system to effectuate benefit adjustments; a website and call center to help beneficiaries use BOND; and removed disincentives in Ticket payment rules for providers not to accept tickets of BOND participants. For more details on the BOND intervention, see Wittenburg et al. (2012).

6 The evaluation's final report for both stages is scheduled to be released in 2018.

7 The *BOND Final Design Report* described the rationale for the offset and presented the demonstration design (Stapleton et al. 2010). The *BOND Evaluation Analysis Plan* provided the detailed plan for evaluation of the BOND innovations, including the methods to estimate impacts for each of outcomes considered in this report and the timeline for reporting outcomes in future reports (Bell et al. 2011). Finally, the *Stage 1 Early Assessment Report* documented and assessed the implementation of the infrastructure to deliver Stage 1 services and examined use of those services in the first six months following random assignment (Wittenburg et al. 2012).

8 As described in the Appendix, we made some modifications to the methodology.

employment outcomes, but they receive less weight than the confirmatory findings in the assessment of the success of the tested treatment.

For reasons identified in previous evaluation reports, we expected that impacts on all outcomes would be small for the eight months covered in this report. As specified in our design report, the impacts on these outcomes might take considerable time to develop given that BOND subjects might not immediately use demonstration services and enroll in the offset (Stapleton et al. 2010). We also know that use of the offset by BOND treatment subjects was very limited through the end of 2011 (Wittenburg et al. 2012).

1.3. Organization of Report

The remainder of this report includes three sections and an Appendix. Section 2 provides background information on the BOND sample and the impact estimation methodology. Section 3 presents the impact findings for the confirmatory and exploratory outcomes for the overall Stage 1 sample and key subgroups. Section 4 includes a brief discussion of the results and implications for future reports. Finally, the Appendix provides a detailed description of the estimation methodology along with additional impact tables for other groups of interest to the evaluation.

2. Background on BOND and Approach to Estimating Impacts

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. Hence, in Stage 1 nearly all current SSDI beneficiaries residing in the demonstration areas were randomly assigned to one of three groups:

- **T1 subjects** are 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** are a control group that continues to receive benefits according to current law. This group initially included a sample of the same size as the T1 group, called the **C1-core subjects.** Following completion of Stage 2 recruitment, the C1 sample was substantially expanded as described below.
- **Stage 2 solicitation pool subjects** are a group from which the demonstration released random replicates for purposes of recruiting volunteers for Stage 2. When Stage 2 recruitment was completed, subjects in the unused random replicates were assigned to C1 (C1-supplement subjects).

The remainder of this section describes the evaluation sample, summarizes findings from the *Stage 1 Early Assessment Report*, considers the anticipated impacts, and discusses the methodology used to estimate impacts.

2.1. Evaluation Sample for Stage 1

Given the expectation that only a small fraction of T1 subjects offered the offset will be likely to use it, the T1 and C1 groups must be very large (tens of thousands of individuals each) in order to detect policy-relevant impacts (Stapleton et al. 2010).[9] Mean impacts across all T1 subjects are expected to be quite small even if mean impacts for those who pursue use of the offset are quite large, because most T1 and C1 subjects are not expected to work at all.

To meet the large sample targets, the BOND sample includes 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. Most of the BOND sample includes primary beneficiaries, who qualified based on their own earnings history. However, the sample also includes auxiliary beneficiaries, who are SSDI beneficiaries who qualify on the basis of connection to a primary beneficiary. Specifically, two auxiliary beneficiary groups, Disabled Adult Children (DAC) and Disabled Widow(er) Beneficiaries (DWBs), are also included in the BOND sample.[10] Additionally, a significant minority of SSDI beneficiaries concurrently

[9] Because of the severity of the SSDI disability eligibility criteria (for examples, the requirement that earnings be below SGA to qualify for benefits), many beneficiaries will not work even with the benefit offset. In Stapleton et al. (2010), we anticipated that benefit offset usage would likely be low based on the current work experiences of SSDI beneficiaries (perhaps less than 5 percent would use the benefit offset and 10 percent would appear to be unlikely).

[10] An adult of any age who first meets the medical eligibility criteria before age 22 becomes eligible for DAC benefits based on a parent's work history when the parent dies or successfully claims Social Security retirement or disability benefits. DWB benefits are based on the earnings history of the deceased spouse, and eligibility is

receive Supplemental Security Income (SSI) benefits (these beneficiaries are defined as "concurrent" beneficiaries as opposed to "SSDI-only" beneficiaries).[11] As described in the *Design Report*, these subgroups are notable because the use of the offset might vary among these different types of SSDI beneficiaries.

The 10 sites were randomly selected from among SSA's 53 area offices throughout the nation using a stratified method designed to ensure that the 10 selected sites represent the universe of area offices. The selected sites include about 20 percent of all current SSDI beneficiaries.

2.1.1. Random Assignment Design

For purposes of random assignment, BOND-eligible beneficiaries were stratified by site and by duration since their first SSDI payment—fewer than 36 months (short duration) and 36 months or longer (long duration). Short-duration beneficiaries were oversampled so that they would constitute one-half of the T1 subjects.[12] The sample was stratified by duration to ensure that enough short-duration beneficiaries would be assigned to the treatment group to support projections of BOND's impacts in a future scenario in which all beneficiaries are subject to the offset when they initially enter SSDI.[13] Based on previous research, it seems likely that the percentage of short-duration subjects who use the offset will be larger than that of long-duration subjects (Stapleton et al. 2010).

The evaluation team implemented Stage 1 random assignment in May 2011. The team randomly selected nearly 80,000 beneficiaries for the T1 group and an equal number for an initial control group (C1-core). We found no statistical differences between the observed characteristics of the T1 and C1-core groups, indicating that random assignment worked as envisioned in the design (Wittenburg et al. 2012). One other group was added to the C1 sample after completion of Stage 2 recruitment (C1-supplement): those BOND-eligible subjects who were not included in the samples that were released for Stage 2 recruitment.[14] The C1 sample is the combination of the C1-core and C1-supplement samples.

---

restricted to widow(er)s age 50 or older who meet the medical eligibility criteria. Some DACs and DWBs are dually eligible, in that they also qualify as primary beneficiaries; for purposes of the evaluation they are not distinguished from other DACs and DWBs.

[11] The SSI program is an income-maintenance program administered by SSA for low-income adults and children. SSI and SSDI use the same disability eligibility determination process to establish disability eligibility. Unlike SSDI, in which beneficiaries qualify based on their work history, SSI applicants must meet income and asset eligibility requirements.

[12] The 36-month requirement was determined based on the beneficiary's status as of June 2011. This date was chosen in order to place the cutoff in the middle of the originally planned three-month mailing effort. The proportion of short-duration beneficiaries within each site is the naturally occurring proportion in the site multiplied by a constant factor (the same for all sites) such that the total number of short-duration beneficiaries in T1 across sites is exactly half of the T1 group (that is, approximately 40,000 beneficiaries).

[13] See Bell et al. (2011) for discussion of why short-duration subjects are expected to behave differently than long-duration subjects.

[14] The samples of BOND-eligible beneficiaries not released for Stage 2 recruitment include all concurrent beneficiaries (not included in Stage 2 by design) and SSDI-only beneficiaries in the Stage 2 solicitation pool not included in the sample replicates that were released for recruitment. These groups were added to C1 after the completion of Stage 2 recruitment.

To maximize the precision of the impact estimates, the analysis uses the full C1 sample. The characteristics of the full C1 sample were not available at the time of our *Stage 1 Early Assessment Report*. For this reason, we examine the baseline equivalence of the T1 and full C1 sample below to confirm that no significant difference emerged as the result of adding the C1-supplement sample.

2.1.2. Sample Sizes

As shown in Exhibit 2-1, the final Stage 1 analysis sample includes a total of 968,713 subjects, spread across T1 (77,115 subjects) and C1 (891,598 subjects). The final sample excludes subjects who died just prior to random assignment, but whose deaths were not identified in administrative records until later. These cases accounted for less than 1 percent of the overall sample. We also have excluded pairs of related beneficiaries who receive disability benefits based on a common primary beneficiary's record if both members of the pair were not assigned to the same Stage 1 group (T1 or C1). A large majority of excluded cases were primary worker beneficiaries assigned to one group with a DAC assigned to the other group.[15] The number excluded in this manner was less than 4 percent of all T1 and C1 subjects. We removed these cases because the behavior of one subject might be influenced directly or indirectly by the fact that different benefit-adjustment rules apply to the earnings of the other subject; to use the language of experimental evaluations, the behavior of both subjects is potentially contaminated by the assignment of the other to a different group. Under a national benefit offset, the same benefit adjustment rules would presumably apply to the earnings of all disabled beneficiaries entitled to benefits via a common primary beneficiary, just as they do under current law today.[16] If members of a pair were both assigned to the same group (either T1 or C1), they were not excluded from the sample. The weights are adjusted to ensure that both the T1 and C1 analysis samples are representative of all those in the national beneficiary population who met BOND eligibility criteria in the month of random assignment.[17] See the Appendix for analytic adjustments that follow from these exclusions.

---

[15] We excluded subjects where any pair was assigned to a different random assignment group in Stage 1 or Stage 2 (e.g., a C1 DAC and a Stage 2 treatment subject). In addition to disabled worker/DAC pairs, we excluded some DAC/DWB and DAC/DAC pairs who were receiving benefits as survivors of a common primary beneficiary. We also found and excluded a small number of beneficiaries who were members of trios and larger family clusters whose members were assigned to different groups.

[16] Although concerns about contamination primarily stem from how assignment of a pair to different groups might affect the behavior of both members of the pair, there is a secondary consideration related to how changes in the earnings of the primary beneficiary might affect the benefits of the DAC. The benefit offset was designed so that increases in the earnings of a primary disabled worker would have no effect on the benefits of auxiliary beneficiaries, including DAC, unless the primary earns so much that the primary benefit is zero, in which case all auxiliary benefits are suspended—an event that seems very unlikely. However, an increase in the earnings of a primary beneficiary might result in an increase in the benefits of a DAC—if the earnings increase is sufficient to increase the Primary Insurance Amount (PIA) of the primary disabled worker.

[17] There is one minor exception to this statement. Groups of three or more BOND subjects who receive benefits under a single primary beneficiary's record (for example, a primary disabled worker with two DACs) are not represented. These beneficiaries represent 0.5 percent of the beneficiary population. See the appendix for details about why this group is not represented in the analysis sample.

Exhibit 2-1. Stage 1 Analysis Sample Composition

Random Assignment Group	Sample Size
T1	77,115
C1	891,598
C1-Core	78,604
C1-Supplement	812,994
Population Size	6,526,888

Source: BOND Operations Data System (BODS).

Notes: The Stage 1 analysis sample excludes subjects initially assigned to the sample but who were later determined to 1) have died prior to assignment, or 2) have a primary beneficiary in common with that of a BOND subject who was assigned to a different BOND group. 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.

2.1.3. Characteristics of Stage 1 Sample

Exhibit 2-2 presents selected characteristics of the weighted Stage 1 analysis sample. Just over half of the beneficiaries are male, and the mean age of the sample was 48 in April 2011 (Exhibit 2-2). Half of T1 subjects have allowances based on mental disorders (31 percent) or musculoskeletal disorders (23 percent). At baseline, the sample's mean SSDI benefit was $995 per month, and only a small share of subjects concurrently received SSI (18 percent). A large majority received benefits only as primary beneficiaries (89 percent); the remainder are DACs and DWBs, including some who were "dually entitled"—entitled as a primary beneficiary based on their own work history and entitled as a DAC or DWB. Finally, 30 percent of BOND subjects were short-duration beneficiaries (i.e., they had received benefits for fewer than 36 months as of random assignment).[18]

Consistent with expectations, we find that baseline characteristics for the weighted T1 sample are statistically equivalent to those for the weighted C1 sample, as well as to those for the weighted C1-core sample. These findings give us a high level of confidence that any statistically significant differences in subsequent outcomes between the T1 and C1 groups will represent real impacts of the benefit offset in the treatment group rather than systematic pre-existing differences between the two groups or their environments.[19]

[18] It is important to note that the unweighted T1 and C1-core samples are approximately evenly split between short- and long-duration beneficiaries. The percentages for the weighted samples in Exhibit 2-2 are unbiased estimates of population percentages.

[19] The findings are consistent with results from a comparison of the T1 sample and C1-core sample prior to exclusion of beneficiary pairs due to possible contamination, as reported in Wittenburg et al. (2012).

Exhibit 2-2. Baseline Characteristics of T1, C1, and C1-Core Subjects Prior to Random Assignment in April 2011, by Site

Baseline Characteristic	Means			Difference	
	T1	C1	C1-Core	T1 vs. C1 Total	T1 vs. C1 Core
Mean age	47.6	47.7	47.7	-0.1	-0.1
Male	51.6	51.5	51.7	0.0	-0.1
Primary Impairment					
Neoplasms	2.6	2.6	2.6	0.1	0.0
Mental disorders	31.2	30.9	30.7	0.2	0.5
Back or other musculoskeletal	22.8	23.1	23.3	-0.3	-0.4
Nervous system disorders	7.2	7.3	7.1	-0.1	0.1
Circulatory system disorders	5.8	5.9	5.9	-0.0	-0.1
Genitourinary system disorders	1.8	1.8	1.8	0.0	0.0
Injuries	4.3	4.2	4.3	0.1	0.0
Respiratory	1.9	2.0	1.9	-0.0	-0.0
Severe visual impairments	1.9	2.1	1.9	-0.1	-0.0
Digestive system	1.6	1.5	1.6	0.1	-0.0
Other impairments	18.7	18.6	18.7	0.1	0.0
Beneficiary Subgroups					
Concurrent	18.2	18.0	17.7	0.2	0.4*
Short-duration	30.2	30.1	30.2	0.1	-0.0
Auxiliary or Other Benefits					
Monthly benefit amount	$997	$996	$999	$1	-$2
Primary beneficiary	88.5	88.8	88.8	-0.2	-0.3
Disabled adult child	13.0	12.8	12.8	0.2	0.2
Disabled widow beneficiary	1.7	1.7	1.7	0.0	0.1
Payee is other than self	18.3	18.6	18.4	-0.3	-0.1
2010 AIME	$1,607	$1,597	$1,602	$10	$5
Site					
Northern New England	3.8	3.9	3.9	-0.0	-0.0
Western New York	15.3	15.3	15.5	-0.1	-0.3
Greater Detroit	12.4	12.5	12.4	-0.1	0.0
Wisconsin	10.4	10.1	10.2	0.3	0.2
Alabama	11.4	11.5	11.5	-0.1	-0.0
South Florida	11.4	11.4	11.6	-0.0	-0.2
Greater Houston	9.6	9.6	9.4	-0.1	0.1
DC Metro	8.2	8.3	8.2	-0.1	0.0
Colorado/Wyoming	5.8	5.8	5.8	-0.0	-0.0
Arizona/Southeast California	11.7	11.5	11.5	0.2	0.3

Source: Analysis of SSA administrative records from the Summary Earnings Record (SER), BODS, Master Beneficiary Record (MBR), and Supplemental Security Record (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. Unweighted sample sizes: T1 = 77,115; C1 = 891,598. AIME is Average Indexed Monthly Earnings.

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

2.2. Synopsis of Findings from the *Stage 1 Early Assessment Report*

To implement BOND, SSA needed to build an administrative infrastructure that was largely external to SSA. As documented in Wittenburg et al. (2012), the BOND implementation team built most of the infrastructure required to communicate with BOND subjects, conduct outreach to relevant entities in the BOND sites, recruit Stage 2 subjects, provide counseling services, and support the processing of earnings and other information—as needed to determine the completion of the TWP and GP and to adjust benefits under the offset. In addition, SSA built an internal component of the infrastructure needed to carry out its adjudicative and fiduciary responsibilities for T1 subjects—primarily to determine TWP and GP status, to adjust benefits under the offset, and to make benefit payments. SSA's existing infrastructure continued to administer the benefits of C1 subjects.

As reported in Wittenburg et al. (2012), SSA and the BOND implementation team did set up the infrastructure envisioned in the original design, but usage of the offset was limited during 2011. Demonstration staff sent an outreach letter to every T1 subjects.[20] Less than 1 percent of those letters were returned, but there was no way to assess the extent to which T1 subjects received the letters, read the material, or understood and believed the content. Only 39 T1 subjects had benefits paid under the offset as of December 2011, much lower than the more than the expected 800 or more offset users. There are several potential explanations for low initial offset usage, some of which reflect the length of time that both treatment subjects and operational entities need to understand how the benefit offset works. T1 subjects might not have received, read, understood, or believed the initial outreach letter. Further, even if they did, they might have realized that their 2011 benefits would eventually be adjusted retroactively under the offset, based on Internal Revenue Service (IRS) records, even if they did not initiate contact with the demonstration. Additionally, our qualitative findings indicated that parts of the infrastructure to provide supports to Stage 1 subjects (e.g., WIC) was not operating as smoothly as was intended during this start-up period.

2.3. Methodology for Estimating Impacts

The impact analysis draws on a limited number of benefit- and earnings-related outcomes that were available in administrative data at the time of this report.[21] The remainder of this section describes the outcome measures used in this report, discusses the hypothesized direction of impacts and their likely size

---

[20] SSA also sent a follow-up letter to T1 subjects that provided details on the offset.

[21] Baseline characteristics of all BOND-eligible subjects were taken from the BOND Operations Data System (BODS); these data were originally drawn from SSA administrative files. Benefit outcomes are from SSA's Master Beneficiary Record (MBR, for SSDI) and Supplemental Security Record (SSR, for SSI). Earnings are from the SSA Master Earnings File (MEF). The MEF contains longitudinal information on wages and self-employment income reported to the IRS, and the records were almost 100 percent complete for calendar year 2011 when SSA extracted them for this report. SSA staff have direct access to MEF data, but contractors do not, because the data are collected by the IRS and therefore subject to IRS access rules. 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 data to the evaluation team. The MEF earnings data are updated annually, with more than 90 percent of the records updated by August of the following calendar year. The MEF data are considered fully updated by the following February. The 2011 earnings data for this report were extracted in November 2012.

in 2011, and provides a summary of the estimation methodology. The section concludes with a discussion of subgroup estimates.

2.3.1. Definitions of Outcomes

Different data sources imply that benefit and earnings impacts are estimated over different periods: benefit impacts are based on monthly administrative data and available for the months from May 2011 through December 2011, while earnings impacts based on annual earnings data are available only for the full calendar year (January 2011 to December 2011).[22] The earnings impacts include a short period before BOND (January through April), though presumably there were no impacts on earnings prior to May. Hence, we assume any impacts from earnings during 2011 represent impacts on earnings after May of that year.

In Bell et al. (2011), we specified many outcomes for the impact analysis, nine of which can be constructed using the data available for this report (Exhibit 2-3). These outcomes include the two confirmatory outcomes: total earnings (annual 2011 earnings in this report) and total SSDI benefits paid (for May to December 2011 only in this report). The exploratory outcomes are also based on earnings 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 amount, two times the BYA amount, and three times the BYA amount) and an indicator for any earnings during 2011. The exploratory benefit outcomes include number of months with SSDI payments, total SSI benefits paid, and number of months with SSI payments.

2.3.2. Expectations for Benefit and Earnings Impacts

The third column of Exhibit 2-3 summarizes the theoretical predictions about the direction of the benefit offset's impacts on these 10 outcomes. As described in Bell et al. (2011), the direction of the predicted impact for most outcomes is ambiguous. This ambiguity arises because the work and earnings incentives created by the benefit offset vary with what the beneficiary's earnings would be under current law. T1 subjects who would have had earnings below or near BYA under current law are expected, on average, to have higher earnings and lower SSDI benefits. Conversely, T1 subjects who would have had earnings well above BYA but below the BOND break-even are expected, on average, to have lower earnings and higher SSDI benefits. Hence, although the benefit offset was designed to increase beneficiary earnings and lower benefits, the theoretical direction of impacts on mean earnings and benefits is ambiguous. There are, however, predicted signs for impacts on five of our seven exploratory outcomes.[23] Theory

---

[22] The reason for using disparate periods is that SSA benefit data are available on a monthly basis, whereas IRS earnings data are available only for the full calendar year.

[23] Theory predicts that the offset will increase both the percentage employed and the percentage of beneficiaries with earnings above BYA, because even those beneficiaries who might reduce their earnings would not reduce them to an amount that is less than BYA. However, it is possible that there will be impacts on earnings well above BYA. For this reason, the direction of impacts on the percentage with earnings above two times BYA and three times BYA is theoretically ambiguous; some T1 subjects might reduce their earnings in response to the benefit offset. The percentage of T1 subjects with earnings above either threshold will not necessarily decline, but it might. 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. Theory also predicts that the impact on SSI benefits paid, applicable only to concurrent beneficiaries, will be negative. Under current law, any concurrent beneficiary engaged in SGA would receive only an SSI payment, if anything, after completing the

predicts positive impacts on employment, earnings above BYA, and months with SSDI payments and negative impacts on SSI benefits and months with SSI payments.

Exhibit 2-3. Definitions of Confirmatory and Exploratory Outcomes and Hypothesized Benefit Offset Impact on Outcomes

	Definition	Sign of Expected Impact
Confirmatory Outcomes		
Total earnings (January–December 2011)	2011 earnings	?
Total SSDI benefit paid	Sum of SSDI benefit payments from May through December 2011; for SSDI workers, this includes benefits for dependent spouses and minor children, but not for DAC[b]; for DAC and DWB, it includes only benefits payable to the DAC or DWB	?
Exploratory Outcomes		
Earnings Outcomes (January–December 2011)[a]		
Employment during year	Any 2011 earnings	+
Earnings above BYA	2011 earnings above $12,000 (non-blind subjects) or $19,680 (blind subjects)	+
Earnings above 2 × BYA	2011 earnings above $24,000 (non-blind subjects) or $39,360 (blind subjects)	?
Earnings above 3 × BYA	2011 earnings above $36,000 (non-blind subjects) or $59,040 (blind subjects)	?
Benefit Outcomes (May–December 2011)		
Number of months with SSDI payments	Number of months with SSDI benefit paid above zero	+
Total SSI benefits paid	Sum of SSI benefit payment amounts from May through December 2011	-
Number of months with SSI payments	Number of months with SSI benefit paid above zero	-

Notes: Bell et al. (2011) provide detailed discussion on the hypothesized impacts of benefit offset.

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

[b] For a description of family benefits, see [http://www.socialsecurity.gov/pubs/10024.html#a0=3]; accessed January 26, 2013.

TWP and GP. 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.

Regardless of the predicted direction of impacts, the size of impacts on any outcome in 2011 is expected to be small for several reasons.

- Most importantly, Stage 1 outreach occurred from May through August 2011, so the offset could affect beneficiary behavior only in four to eight months of 2011.
- It is likely to take some time before any response is translated into a change in earnings—it takes time to find a job or even to increase earnings at an existing job. Benefit changes could take even longer to emerge because earnings increases do not affect benefits under the offset until the beneficiary has completed the nine TWP months and three GP months. Only 10 percent of T1 subjects had completed their TWP as of October 2011 (Wittenburg et al. 2012).
- Changes in benefits paid for T1 subjects would be further delayed by delays in the review of TWP and GP status and in the processing of benefit adjustments. The "benefits paid" variable reflects the benefits SSA actually paid the beneficiary during the period. Retroactive adjustments to benefits based on post-2011 reviews of earnings during this period will be reflected in the benefits paid in later years.
- The reaction of T1 subjects might have been significantly muted by limited information about, understanding of, or trust in the opportunity offered by, the offset (see Wittenburg et al. 2012).
- Finally, it is possible that the recession dampened or delayed the impacts of the benefit offset on employment and earnings relative to what they would have been in a stronger labor market. It seems likely that the weak economy reduced the employment and earnings of both T1 and C1 subjects. This dampening effect is not necessarily larger for T1 subjects than for C1 subjects. However, findings from previous welfare to work and job training demonstrations indicate that poor economic conditions could dampen impacts, especially on earnings (Bloom et al. 2003; Greenberg et al. 2003; Heinrich 2002).

2.3.3. Impact Estimation and Testing Methodology

The goal of the Stage 1 BOND experiment is to make inferences about what the impact of the benefit offset would be if applied to all SSDI beneficiaries in the nation meeting the BOND eligibility criteria as of May 2011. The statistical design of the BOND sample supports the production of unbiased point estimates and standard errors (SEs) for this population. The SEs reflect both random variation associated with the selection of the BOND sites as well as the random variation associated with assignment of subjects in those sites to T1 and C1.[24] 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 used are "intent to treat" estimates. They estimate the mean impact of the applicability of the benefit offset rules to the earnings of all T1 subjects, including the large majority who would not have any earnings under current law or the offset as well as those with earnings who fail to learn about, understand, or trust the offset. We expect that the offset rules will affect the earnings and

[24] The point estimates reported here may also be interpreted as unbiased estimates of impacts for BOND-eligible beneficiaries in the BOND sites, conditional on the sites actually selected. However, the SEs reported are somewhat larger than the corresponding conditional SEs, as the conditional SEs would reflect variation only due to random assignment of BOND subjects in the BOND sites to the T1 and C1 groups.

benefits of only a small share of treatment subjects, so mean intent to treat impacts will be small. This expectation is the reason that the T1 sample is so large.

The impact estimation methodology used in this report differs from the planned methodology presented in Bell et al. (2011), and is both more stable and more computationally efficient than the original approach.[25] The method is described in detail in the Appendix to this report. The method compares mean outcomes for the T1 group to mean outcomes for the C1 group that have been weighted for differences in sampling rates across sampling strata and adjusted for the effects of small differences in baseline characteristics. The adjustments correct for any chance differences in baseline characteristics between the two groups and also reduce the SEs.

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

Results of multiple tests of this sort can be misleading, because the more such tests are conducted, the more likely it is that at least one result will reject its null hypothesis even if all null hypotheses are true (i.e., there is no true impact of the intervention on any outcome—overall or for any subgroup). Thus, if all null hypotheses tested are true, and multiple individual tests are conducted using the 5 percent significance level, the probability of finding at least one significant impact will be greater than 5 percent.

To address the multiple comparisons problem, we first selected two outcomes to be the "confirmatory" outcomes for BOND, based on theory and policy interest alone (see Bell et al. 2012): total earnings and total SSDI benefits paid. The evaluation is using estimates of impacts on means for these outcomes to confirm that the benefit offset has impacts on earnings and benefits. We then chose a method to adjust test statistics for these outcomes that addresses the multiple comparison issue described above. If we performed the two individual tests for these outcomes without any adjustment, then the probability of rejecting the null hypothesis for at least one outcome if the null hypothesis is true for both outcomes would exceed the specified significance level for each individual test. Instead, we adjust the test statistics for each of the two outcomes in a manner that reduces the probability of rejecting the null hypothesis of no impact on either confirmatory outcome if the null hypothesis is true to the specified significance level.[26]

The same adjustment is not applied to tests for the exploratory outcomes. These tests are exploratory because their purpose is to explore the possibility of other impacts, rather than to confirm that the benefit

---

[25] We departed from the planned method described in Bell et al. (2011) in order to reduce the considerable computational burden of producing estimates from such large samples. First, we added a data reduction step in order to speed computation. As discussed in section A.1 of the Appendix, this step is also appealing from a statistical perspective. Second, we changed the estimation model from hierarchical linear modeling (HLM) to survey methods (as implemented in SAS' PROC SURVEYREG) to ensure computational stability (i.e., to avoid a potential problem with model convergence). Additional explanation and full details of our revised approach appear in the Appendix.

[26] 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 the Appendix. See Schochet (2008) for further discussion of the multiple-comparisons problem.

offset had impacts. It must be recognized, however, that the probability of finding at least some statistically significant impacts in these exploratory tests even if all true impacts are zero is higher than the significance level for each test—likely considerably higher given the number of tests performed. This undermines the evidentiary value of any significant result. Hence, readers are advised to give less weight to any individual significant result from an exploratory test than they would to an equally significant result from a confirmatory test. It is appropriate to put more weight on a result from an exploratory test that is statistically strong (for example, is significant at the 1 percent level); that is one result in a consistent pattern of results (for example, is replicated for multiple mutually exclusive subgroups); and/or has a sign that is consistent with an unambiguous theoretical prediction (that is, those unambiguous predictions indicated in Exhibit 2-3).

2.3.4. Impact Estimation for Subgroups Defined by Duration of Benefit Receipt and SSDI Benefit Status

We present impacts for the overall Stage 1 BOND sample and for two subgroups defined by duration of benefit receipt and SSI benefit status. We treat all subgroup analyses as exploratory.

Short-duration SSDI beneficiaries are an important subgroup because they provide the evaluation with the opportunity to learn how beneficiaries who recently entered the rolls will respond to the benefit offset. Given that these beneficiaries were attached to the labor force relatively recently, their response to the offset might be quite different than the response of those who have been on the rolls for many years (long-duration SSDI beneficiaries). If so, the long-run impacts—when all T1 subjects have had the opportunity to use the offset since their first day on the rolls—might be substantially different from the impacts during the first years after implementation. Hence, tracking the outcomes of short-duration beneficiaries will improve our understanding of the long-term impacts of a national program.

The second subgroup is for concurrent beneficiaries. As discussed earlier, this distinction is of interest because the interaction between SSI benefits and SSDI benefits under the offset is such that the value of the SSDI offset to a concurrent beneficiary is smaller than is the value of the offset to a T1 SSDI-only subject with a comparable C1 SSDI-only subject.

3. Findings

This section presents impact estimates for the two confirmatory outcomes and seven exploratory outcomes summarized in Section 2 for the full Stage 1 impact sample.[27] We first present impact estimates for the full Stage 1 BOND sample and then summarize findings for subgroups defined by duration of benefit receipt and SSI status in the month prior to random assignment. For each outcome, we show the impact estimate, measured as the difference between the weighted T1 group mean and weighted C1 group mean after statistical adjustments to the latter for differences in observed characteristics (see Section 2). When comparing outcome means between groups, we cite weighted means for subjects that have been adjusted via regression to the mean baseline characteristics of the T1 subjects.

We report statistical significance at the 1, 5, and 10 percent levels for all impact estimates. The only confirmatory outcomes, which include the multiple comparisons adjustments outlined in Section 2, are total earnings and total SSDI benefits paid for all Stage 1 subjects. The remaining outcomes and all of the outcomes for the subgroup analysis (including total earnings and total SSDI benefits paid) are exploratory; hence, statistical tests for impacts on these outcomes do not include a multiple comparisons adjustment. We describe impact estimates that are statistically significant at a 1 percent level as "strong evidence," 5 percent level as "evidence," and 10 percent level as "marginal evidence." We term as insignificant any difference that is not significant at even the 10 percent level.

We are able to detect very small impacts for several outcomes, especially benefits paid, which reflects the size of our sample and the strong predictive power of our regression adjustment models for these outcomes. For example, our model includes benefits paid just prior to random assignment, which is, not surprisingly, highly predictive of benefits paid following random assignment given that most SSDI beneficiaries have the same beneficiary amount in each month. To assess the substantive importance of any significant impact estimate, we express it as a percentage of the corresponding control group mean; the latter is an unbiased estimate of what the mean outcome for the treatment group would be in the absence of the benefit offset. As will be seen, some significant impacts on benefits paid are very small as a percentage of the adjusted control group mean.

3.1. Full Stage 1 Treatment Group

Exhibit 3-1 presents the estimates of impacts on earnings and benefit outcomes for the full Stage 1 BOND treatment group. As described in Section 2, total earnings (January–December 2011) and total benefits paid (May–December 2011) are the confirmatory outcomes. All remaining earnings and benefit outcomes are exploratory, so their statistical tests reflect no such adjustments.

3.1.1. Confirmatory Impacts: No Earnings Impacts, Very Small Increase in Benefits Paid

The benefit offset had no statistically significant impact on total earnings in 2011. Mean total earnings for C1 subjects were low for calendar year 2011 ($1,204), reflecting the fact that most C1 and T1 subjects had no earnings in 2011.

[27] The Appendix examines the sensitivity of the findings to use of the C1-core sample alone and to inclusion of all BOND-eligible beneficiaries who are members of families of beneficiaries. In each case the results are not substantively different from those presented in this section.

There was a very small positive and marginally significant impact on total SSDI benefits paid for the May–December 2011 period. The estimated impact on mean benefits paid was $23, representing a $3 increase in benefits paid per month ($23 divided by eight months), or just a 0.3 percent of mean total SSDI benefits paid to C1 subjects for May–December 2011 ($7,508). This finding does not seem important from the perspective of SSDI program costs because of its small magnitude and only marginal significance.[28] It is also important to recognize that the small mean impact for all T1 subjects might reflect a much larger mean impact for the small subgroup of T1 subjects who benefited from the offset in 2011. Further, as discussed in the final section of the report, the estimated impacts on benefits paid in 2011 do not reflect impacts on retroactive benefit adjustments for 2011 made after 2011.

3.1.2. Exploratory Impacts: No Impacts on Any Outcomes

There were no statistically significant impacts for any of the four exploratory earnings outcomes. Just over 16 percent of subjects in each group had at least some earnings in 2011, including 2.4 percent with earnings above BYA, 1.0 percent with earnings above twice BYA, and 0.5 percent with earnings above three times BYA.

There were also no statistically significant impacts for the three exploratory benefit outcomes. The mean number of months with SSDI payments was 7.5 (out of a possible 8.0 months). The lack of a significant positive impact for this outcome underscores the weak nature of the impact finding for mean SSDI benefits paid noted above. Theory predicts that the impact of the benefit offset on an individual's benefits will be positive only if the beneficiary would have received no payment under current law, but receives a partial payment under the benefit offset. Hence, we would expect a positive impact on benefits paid only if there is also a positive impact on the number of months with benefit payments. If there was a positive impact on the latter, it was too small to be detected.

For T1 and C1, the mean total SSI payment was just over $340 over eight months and the mean number of months with an SSI payment was 1.4 months. The mean of total SSI benefits paid was small in comparison to the mean of total SSDI benefits paid ($7,508 for C1 subjects), reflecting the fact that only a small minority of Stage 1 subjects received SSI benefits in 2011.

[28] We also examined the distribution of SSDI benefits paid to assess whether outliers could be driving any of these small differences. We found a number of outlier values for benefits, though we did not make any adjustments to these outcomes, in part because our empirical model without any outlier adjustment produced very precise standard errors. The outliers are problematic only in that they increase standard errors, making it more difficult to detect small impacts. SSA's investigation of the outliers found no evidence that they reflect data entry errors. Outlier values for benefits occur because SSA sometimes makes retroactive benefit payments, especially for new SSDI beneficiaries. We did a similar investigation for earnings and found a few cases of large earnings. Outlier values for earnings can occur for many reasons, including large payouts by employers.

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

	T1 Mean	C1 Mean	Impact Estimate
Earnings Outcomes (January–December 2011)			
Total earnings **(confirmatory)**	$1,195	$1,204	-$9 ($25)
Employment during year	16.15%	16.03%	0.13 (0.10)
Earnings above BYA [a]	2.43%	2.41%	0.02 (0.12)
Earnings above 2x BYA	0.95%	0.97%	-0.03 (0.05)
Earnings above 3 x BYA	0.53%	0.53%	0.00 (0.03)
Benefit Outcomes (May–December 2011)			
Total SSDI benefits paid **(confirmatory)**	$7,531	$7,508	$23* ($10)
Number of months with SSDI payments	7.49	7.49	0.00 (<0.01)
Total SSI benefits paid	$340	$342	-$2 ($5)
Number of months with SSI payments	1.37	1.38	-0.00 (<0.01)

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

Notes: Weights are used to ensure that the BOND subjects who met analysis criteria 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. See Chapter 3 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 the Appendix for more details on the statistical tests and adjustments to the *p*-values). Tests for impacts on all other outcomes (exploratory outcomes) were conducted independently, without multiple comparison adjustments.

*/**/*** Impact estimate is significantly different from zero at the .10/.05/.01 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) in the month prior to random assignment. The outcomes are the same as those in Exhibit 3-1, but stratified by subgroup. For the reasons outlined in Section 2, we consider all subgroup estimates as exploratory outcomes; hence, we did 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 reflect population differences for the subgroups under current law.[29] We expect differences across each pair of subgroups (see Section 2), which is an important motivation for

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

the subgroup analysis. 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: Limited Evidence of Impacts

In this section, we first show that mean outcomes for the short- and long-duration subgroups are very different under current law, as anticipated. We then show that there is no evidence of differences in impacts across subgroups, reflecting very little evidence of impacts within each subgroup.

The means of total earnings and SSDI benefits paid for the short- and long-duration C1 subgroups in the 2011 follow-up period illustrate the different outcomes for these subgroups under current law. Consistent with our expectations in designing these subgroups (Bell et al. 2011), in 2011 short-duration C1 subjects had higher earnings ($1,337 versus $1,146) and were more likely to be employed (16.7 versus 15.7 percent) than long-duration subjects. The findings are consistent with past research demonstrating that recent entrants are more likely to have earnings than those who have been on the rolls for a longer period (Liu and Stapleton 2011). Additionally, short-duration subjects had higher SSDI benefit payments ($8,300 versus $7,198), which likely reflects the way that SSA indexes pre-SSDI earnings when calculating benefit amounts. More specifically, SSA uses an average wage index (AWI) to inflate past earnings prior to calculating the initial benefit amount; after that, SSA adjust benefits for inflation each year using a Consumer Price Index (CPI). As the AWI typically increases faster than the CPI, mean benefits for new awardees typically increase every year after adjustment for overall inflation. Total SSI payments are also higher for the short-duration group ($376 versus $327), likely reflecting the differences in the pathways to SSI entry for those receiving SSI benefits in these two subgroups. The higher prevalence of SSI and lower mean SSI benefits for long-duration subjects likely reflects relatively longer periods since disability onset for these subjects in comparison to short-duration subjects. Although some beneficiaries in both groups enter SSI before or at the same time they enter SSDI, others enter SSDI first and enter SSI only after their other income and resources fall to levels that are both below their respective thresholds for the SSI means tests. Those in the long-duration group have had more time to spend down their resources.

As shown in column 7 of Exhibit 3-2, there were no outcomes for which impacts differed significantly between short- and long-duration beneficiaries. As shown in columns 3 and 6, there were no statistically significant impacts for any of the five earnings-related outcomes within either subgroup and only a small impact for one of the four benefit outcomes within one subgroup: a marginally significant, positive impact on SSDI benefits paid to long-duration subjects. The point estimate was very small ($18) and represents less than 0.3 percent of the C1 mean ($7,180). This result is consistent with the marginally significant positive impact on mean SSDI benefits paid to all T1 subjects, discussed previously. Here too, however, the small size of the point estimate and the lack of a significant positive impact on months with benefit payments suggest that this finding is not substantively important.

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

	Short-Duration			Long-Duration			Difference in Impact
	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	T1 Mean (4)	C1 Mean (5)	Impact Estimate (6)	(7)
Earnings Outcomes (January–December 2011)							
Total earnings	$1,300	$1,337	-$37 ($40)	$1,149	$1,146	$3 ($29)	-$40 ($49)
Employment during year	16.80%	16.73%	0.06 (0.23)	15.88%	15.72%	0.15 (0.14)	-0.09 (0.27)
Earnings above BYA [a]	2.69%	2.75%	-0.06 (0.10)	2.32%	2.27%	0.05 (0.13)	-0.11 (0.16)
Earnings above 2x BYA	1.11%	1.20%	-0.09 (0.07)	0.88%	0.88%	0.00 (0.05)	-0.09 (0.09)
Earnings above 3x BYA	0.68%	0.70%	-0.02 (0.08)	0.47%	0.46%	0.01 (0.03)	0.01 (0.09)
Benefit Outcomes (May–December 2011)							
Total SSDI benefits paid	$8,300	$8,270	$30 ($19)	$7,198	$7,180	$18* ($9)	$12 ($21)
Number of months with SSDI payments	7.57	7.57	0.00 (0.01)	7.46	7.46	0.00 (0.01)	0.00 (0.01)
Total SSI benefits paid	$368	$376	-$8 ($5)	$328	$327	$1 ($6)	-$9 ($8)
Number of months with SSI payments	1.09	1.09	0.00 (0.00)	1.50	1.50	-0.01 (0.01)	0.01 (0.01)

Source: SSA administrative records, from the MEF, BODS, MBR, 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: short-duration: T1 = 38,669; short-duration C1 = 209,790; long-duration T1 = 38,446; long-duration C1 = 681,808. See Chapter 3 for variable definitions. Impact estimates are regression-adjusted. Benefit impacts are 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/.05/.01 levels, respectively, using a two-tailed t-test.

3.2.2. SSI Benefit Status: No Evidence of Differential Impacts

In this section, we first show that mean outcomes for the SSDI-only and concurrent subgroups are very different under current law, as anticipated. We then show that there is no evidence of differences in impacts across subgroups, reflecting very little evidence of impacts within each subgroup.

The levels of total earnings and SSDI benefits paid for SSDI-only and concurrent beneficiaries in C1 illustrate the different economic experiences of these two subgroups under current law (Exhibit 3-3). Relative to concurrent subjects, SSDI-only subjects had higher mean SSDI benefit payments ($8,356 versus $3,696) and higher mean earnings ($1,308 versus $735), which is consistent with expectations because SSDI-only beneficiaries generally have more substantial work histories than concurrent beneficiaries. Reflecting their more substantial work histories, SSDI-only subjects on average are older than concurrent subjects, have more income from other sources, have higher levels of education, and have acquired more skills through experience.[30] Age and income likely reduce the probability that a beneficiary works (other things constant), whereas education likely increases the earnings of those who do work. The percentage employed in 2011 for both groups was approximately the same (about 16 percent), so the large difference in mean earnings indicates that SSDI-only beneficiaries who worked in 2011 earned much more than concurrent beneficiaries who worked. Given that concurrent subjects are identified based on SSI payments at random assignment, it is no surprise that concurrent subjects had substantially higher mean SSI payments in the eight months following random assignment than did SSDI-only beneficiaries ($1,713 versus $37). The fact that subjects in the SSDI-only group received SSI benefits after random assignment might reflect sufficient declines in assets or income from other sources to satisfy the SSI means test.

As shown in column 7 of Exhibit 3-3, there were no outcomes for which impact estimates differed significantly between SSDI-only and concurrent beneficiaries. There also were no significant impact estimates for any of the earnings or benefit outcomes for concurrent subjects (column 6). Two impact estimates for the SSDI-only group are significant, but very small: a $20 increase in mean SSDI benefits paid and a $4 decrease in mean SSI benefits (column 3). Both estimates are very small relative to the C1 group's level of SSDI benefits paid ($8,356). The small but significant mean impact for SSDI benefits paid mirrors the findings reported earlier for all beneficiaries and for long-term beneficiaries. Further, as with the earlier benefits paid estimates, these impacts are not corroborated by positive significant impacts on months of SSDI benefit receipt.

[30] See Wright et al. (2011) for descriptive statistics on SSDI-only and concurrent beneficiaries from the 2010 National Beneficiary Survey.

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 Outcomes (January-December 2011)							
Total earnings	$1,302	$1,308	-$6 ($31)	$713	$735	-$22 ($21)	$16 ($37)
Employment during year	16.30%	16.13%	0.17 (0.12)	15.50%	15.57%	-0.06 (0.27)	0.23 (0.30)
Earnings above BYA [a]	2.71%	2.66%	0.06 (0.13)	1.16%	1.31%	-0.15 (0.10)	0.21 (0.16)
Earnings above 2x BYA	1.12%	1.14%	-0.02 (0.07)	0.17%	0.22%	-0.06 (0.05)	0.04 (0.09)
Earnings above 3x BYA	0.63%	0.63%	0.00 (0.04)	0.07%	0.06%	0.01 (0.03)	-0.01 (0.05)
Benefit Outcomes (May-December 2011)							
Total SSDI benefits paid	$8,376	$8,356	$20* ($10)	$3,726	$3,695	$31 ($18)	-$11 ($21)
Number of months with SSDI payments	7.54	7.54	-0.00 (0.00)	7.29	7.26	0.03 (0.02)	-0.03 (0.02)
Total SSI benefits paid	$33	$37	-$4 ** ($2)	$1,723	$1,714	$10 ($23)	-$14 ($23)
Number of months with SSI payments	0.07	0.07	-0.00 (0.00)	7.25	7.28	-0.02 (0.02)	0.02 (0.02)

Source: SSA administrative records, from the MEF, BODS, MBR, 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: SSDI-only: T1 = 64,709; SSDI-only: C1 = 694,270; concurrent: T1 = 12,406; concurrent: C1 = 197,328. See Chapter 3 for variable definitions. Impact estimates are regression-adjusted. Benefit impacts are 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/.05/.01 levels, respectively, using a two-tailed t-test.

4. Discussion

In the first eight months of the Stage 1 demonstration (May–December 2011), the estimated impacts of the benefit offset on benefit and earnings outcomes were statistically insignificant or very small for the overall T1 group. In Exhibit 4-1, we summarize the impact findings and compare them to the theoretically expected sign of the impacts outlined in Exhibit 2-3.

For the two confirmatory outcomes, there was no significant impact for total earnings. There was a positive, marginally significant impact on SSDI benefits paid. The impact estimate is very small, however—equivalent to $3 per month, representing less than 0.3 percent of the adjusted total SSDI benefit paid to C1 subjects. Further, it is not corroborated by a positive significant impact estimate for months with SSDI payments. Estimates of impacts for all exploratory outcomes over the full group were all insignificant.

There were no differential impacts across subgroups, defined by duration of SSDI receipt and SSI benefit status at random assignment, reflecting the fact that most impacts within subgroups are not significant and others are very small. All of these estimates are treated as exploratory.

Exhibit 4-1. Summary of Impact Findings

	Sign of Expected Impact	Impact Findings	C1 Mean (Full Sample)
Confirmatory Outcomes			
Total earnings (January–December 2011)	?	No impacts	$1,204
Total SSDI benefits paid (May–December 2011)	?	Positive impact ($23)[a]	$7,508
Exploratory Outcomes			
Earnings Outcomes (January–December 2011)			
Employment during year	+	No impacts	16.03%
Earnings above BYA	+	No impacts	2.41%
Earnings above 2x BYA	?	No impacts	0.97%
Earnings above 3x BYA	?	No impacts	0.53%
Benefit Outcomes (May–December 2011)			
Number of months with SSDI payments	+	No impacts	7.49
Total SSI benefits paid	-	No impacts[b]	$342
Number of months with SSI payments	-	No impacts	1.38

Note: All estimates in the table are for the full sample. See the footnotes for information on significant subgroup estimates.

[a] In exploratory analysis, we also found significant positive impacts on total SSDI benefits paid for the long-duration and SSDI-only subgroups ($18 and $20, respectively).

[b] In additional exploratory analysis, we found a significant negative impact on total SSI benefits for the SSDI-only subgroup (-$4).

The lack of substantial impacts is consistent with expectations based on factors identified in earlier evaluation reports (Stapleton et al. 2010; Wittenburg et al. 2012):

- **Short time period covered in the follow-up.** The time period covered in this report includes only the first eight months following random assignment. Consequently, T1 subjects had relatively little time to adjust their employment and earnings behavior between random assignment (May 2011) and the period when the impacts included in this report were measured (May through December 2011). Further, because notifications were mailed in batches from May through August 2011, most T1 subjects did not learn about the offset immediately after random assignment; some had as few as four months to respond.
- **Eligibility to use the offset was limited.** To use the offset, T1 subjects must have completed their TWP and GP, but only a minority (approximately 10 percent) had completed their TWP by the end of 2011 (Wittenburg et al. 2012).
- **Weak labor market.** On the heels of the most severe recession since the Great Depression, the labor market remained very weak in 2011. The weak labor market potentially dampened the impact of the offset on employment responses. Previous research has documented the negative relationship between weak labor markets and estimates of impacts for employment and training interventions (Bloom et al. 2003; Greenberg et al. 2003; Heinrich 2002).
- **Limited information about the offset.** Although the demonstration mailed outreach letters to all T1 subjects, some subjects might not have received, read, understood, or trusted it. Additionally, they could not necessarily count on trusted sources of information, such as disability organizations or service providers, to corroborate or help them understand information provided by the demonstration. Although reliable information was available from the demonstration, that information would be of little use to a beneficiary who did not know about it, did not know how to access it, or did not trust it.
- **Retroactive adjustment of benefits.** Even after the beneficiary has completed the TWP and GP, it usually takes considerable time for SSA to adjust benefits under the offset. Wittenburg et al. (2012) reported that SSA had adjusted the benefits of only 39 T1 subjects under the offset as of the end of 2011 and projected that SSA would eventually adjust the 2011 benefits of 800 or more T1 subjects. Any such retroactive adjustments to 2011 benefits are not reflected in the benefits paid variable, which represents the amount SSA actually paid to the subjects during the period. Adjustment delays also apply to comparable C1 subjects, but the size of the adjustments are likely different because of the difference between current law and benefit offset rules.

As indicated earlier, we do not consider the significant but very small positive estimate for the impact on mean SSDI benefits paid to be substantively important. It is also difficult to assess how the impact on benefits paid will change in the future given the theoretical predictions outlined for each outcome (see Exhibit 4-1).

The retroactive adjustment described in the last bulleted item could substantively affect the estimates for SSDI benefits paid in both 2012 and 2013. As of the end of 2012, according to BODS data (not shown), SSA had applied the offset to the benefits of 295 T1 subjects, although not necessarily for 2011 earnings. A large, but unknown, number of retroactive adjustments for 2011 were still pending. Because there is no

evidence of impacts on earnings in 2011, we expect that most of these adjustments will be made to the benefits of T1 subjects who would have lost all of their benefits under current law during at least some of the last eight months of 2011—also retroactively. The direction and size of the effects in these years will depend on how rapidly retroactive adjustments occur, the extent to which SSA is able to recover overpayments for the 2011 period, differences in the speed of retroactive adjustments for the T1 and C1 groups, and any differential response of T1 and C1 earnings to these adjustments.[31]

Significant impacts might emerge in future years for earnings outcomes, but theory implies that the expected sign of impacts for mean earnings is ambiguous. As the demonstration matures, the direction of impacts on mean earnings (if any) should become more apparent as T1 subjects presumably gain a better understanding of BOND, as SSA makes retroactive adjustments to their benefits, and as more subjects become eligible for the offset by completing their TWP and GP. It might take longer to establish the direction of the long-term impacts on benefits than on earnings, because changes in earnings affect benefits paid only after completion of the TWP and GP, plus any additional months needed for SSA to determine that these periods are completed and to adjust benefits accordingly.

Several other important factors might also affect the course of future impact estimates, some external to BOND and others internal. Externally, the strength of the economic recovery after 2011 could influence impacts. Internally, in mid-2012 BOND initiated follow-up outreach efforts designed to ensure that T1 subjects adequately understand their opportunity to use the benefit offset. Early beneficiary responses to these efforts suggest that more T1 subjects will take advantage of the offset as a result. In addition, the processing of T1 reconciliations for 2011 in January 2013 is likely to increase awareness of the opportunity available under BOND among T1 subjects who demonstrated the capacity to earn more than BYA in 2011 but had not previously sought to have their benefits adjusted.

Future reports will document the trajectory of impacts on the same annual earnings and benefit outcomes through 2017. Five planned reports will document BOND impacts and other outcomes for Stage 1. Additionally, two synthesis reports will document findings from Stages 1 and 2 (see Bell et al. [2011] for more details). Together, these seven reports will update impacts on the outcomes presented here and include additional evaluation findings. The other findings for Stage 1 include estimates of impacts for an expanded set of outcomes, such as TWP completion, overpayments, use of Ticket to Work, and household income; findings from the process study on the demonstration's implementation; details on participation in the offset; and, after all impact estimates are available, cost-benefit estimates. As with this report, quantitative analyses for future reports will rely heavily on administrative records, but they will also incorporate information from a survey of 10,000 T1 and C1 subjects, which is to be conducted approximately 36 months after enrollment.

[31] In future reports, we will be able to estimate the impact of the benefit offset on mean "benefits due" for T1 subjects in 2011—the benefit amounts that were due in 2011 after all retroactive adjustments have been made. That estimate will provide an indication of the effect of delays in benefit adjustments on impacts for benefits paid in later years.

References

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

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

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

Heinrich, Carolyn J. "Outcomes–Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness." *Public Administration Review*, vol. 62, no. 6, 2002, pp. 712–725.

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

Livermore, Gina, 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." Submitted to the Social Security Administration. Washington, DC: Mathematica Policy Research, 2009.

Mamun, Arif, Paul O'Leary, David Wittenburg, and Jesse Gregory. "Employment Among Social Security Disability Program Beneficiaries: 1996–2007." *Social Security Bulletin*, vol. 71, no. 3, 2011, pp. 11–34.

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

Stapleton, David C., Stephen H. Bell, David C. Wittenburg, Brian Sokol, and Debi McInnis. "BOND Implementation and Evaluation: BOND Final Design Report." Submitted to the Social Security Administration, Office of Program Development & Research. Cambridge, MA: Abt Associates, December 2010.

Westfall, Peter H., Randall Tobias, and Russell D. Wolfinger. *Multiple Comparisons and Multiple Tests Using SAS*. Cary, NC: SAS Institute, 2011.

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

Wittenburg, David, David Stapleton, Michelle Derr, Denise W. Hoffman, and David R. Mann. "BOND Stage 1 Early Assessment Report. Final Report Submitted to the Social Security Administration." Cambridge, MA: Abt Associates, May 2012.

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: Detailed Summary of Methodological Approach and Additional Impact Estimates for C1-Core Group

This appendix describes the method used to estimate the impacts presented in this report. Since the development of the initial model in Bell et al. (2011), we used simulations to gauge run-times for alternative models. Run-time is a major consideration given we will use the same method to estimate impacts for a large number of outcomes using both survey and administrative data in future reports. In testing the method specified in Bell et al. (2011) using simulated data, we found the run times had the potential to be very long, in part because of the large number of sample members and in part because of potential difficulty reaching convergence. We developed an alternative estimation procedure that results in a more efficient process for estimating impacts for the demonstration with virtually no change in the parameter estimates or estimated standard errors.[32] For this reason, we decided to use this new procedure to generate impact estimates for this and all future Stage 1 impact reports.

We also test the sensitivity of our impact findings for the full Stage 1 sample (Exhibit 3-1) to alternative sample specifications. We first rerun our estimates including all beneficiaries who are members of beneficiary families (that is, without adjustment for contamination). Substantive differences between these results and those reported earlier might arise because random assignment of family members to different groups affects behavior of each member in ways that differ from the effect that would occur if the other member(s) were assigned to the same group. Substantive differences might also arise because these estimates include BOND-eligible members of all families with three or such members, whereas all such beneficiaries are excluded from the earlier estimates.

We also estimate the models for all subjects using just the C1-core group, rather than the full C1 group. We produced these estimates to verify that inclusion of C1-supplement subjects, weighted to reflect

[32] In Bell et al. (2011), we presented a hierarchical linear model (HLM) that could be used to estimate benefit offset impacts in both Stage 1 and Stage 2. The model that included baseline covariates (for variance reduction) and analysis weights (to make impact estimates nationally representative) and takes account of the potential variability of BOND's impact from place to place when testing for significant demonstration effects. The revised estimation procedure used in this report and presented in Section A.1 shares all of these features while being more computationally stable (through a change from HLM to a survey methods model) and more computationally efficient (through the use of a data reduction step) . Tests of the original planned HLM method with simulated data indicated that the estimation procedure might have difficulty converging. In particular, the relatively low number of BOND sites (10 sites) made the estimation of the cross-site variance in impacts problematic. In order to ensure that the estimation did not encounter a convergence problem, we changed the basic methodology from HLM to survey methods, as implemented in SAS's PROC SURVEYREG. The survey approach to standard error estimation incorporates the same assumptions about error correlation as HLM without requiring estimation of the non-essential parameter for cross-site impact variance, thereby avoiding a potential difficulty in convergence. There is no loss of precision or validity of national effect estimates as a result of the change in methodology. The only disadvantage of the change in methodology is that the revised approach does not estimate the variability of impacts across the country. Instead, the revised approach focuses on estimating the average national effect of the program if it were to be implemented nationwide. The originally proposed methodology would (if feasible) have permitted us to also predict average variability of effects across area offices. This variability is not of substantive policy interest because no consideration is being given to permanently implementing the program on a selective basis across area offices.

sampling probabilities, does not have a material impact on the results other than to increase precision. As outlined in Bell et al. (2011), the value of this test arises from the greater transparency and conceptual symmetry of the T1-versus-C1 core comparison.

In what follows, we provide details on the econometric model that will be the basis for all impact estimates in the Stage 1 BOND evaluation. Specifically, we describe the estimation procedure, the multiple comparisons procedure, covariates included in the estimation model, and the construction of analysis weights. The appendix concludes with the findings from the sensitivity tests.

A.1. Estimation Procedure

We start 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:

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

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:

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

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

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

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. normally 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 nonnormal distributions of earnings and binary outcomes. To accomplish 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 results in 4,000 random groups (10 sites × 2 assignment groups × 200 random groups). Within each random group, the average residual[33] 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.

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* residuals 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 following estimation model used in this report:

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

$$\bar{R}_{kaj} = \frac{1}{\sum_{m=1}^{n_{kaj}} w_m} \sum_{m=1}^{n_{kaj}} w_m (y_m - \hat{y}_m)$$

where , 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 = the model intercept,

β_1 = the overall impact of the T1 treatment (versus the no treatment of the C1 group), and

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

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

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

The estimation of Equation (3) incorporates the weights of the random groups in order to produce nationally representative results. We estimate Equation (3) using the PROC SURVEYREG procedure in the SAS software package.[34]

A.2. 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, i.e., 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, the hypothesis tests are separated into "confirmatory" and "exploratory" tests, as specified in Bell et al. (2011), prior to the conduct of the impact analysis. Only the two most important outcomes from the evaluation—total earnings and total SSDI benefits paid—are included in the confirmatory group. [35] 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.

---

[34] 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 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.

[35] 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.

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.[36] 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 stepdown 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 stepdown 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 t-tests of impacts on outcomes A and B. These are the "raw," unadjusted *p*-values for each outcome.

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

OUTCOME1, OUTCOME2 = 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.

$p_{OUTCOME1}^{raw}, p_{OUTCOME2}^{raw}$ = raw *p*-values in order from smallest to largest.

We then form some large number *R* (such as 20,000) permutation replicates. With each replicate sample, we run impact regressions for the two outcomes, producing two *p*-values.

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

$$p_{OUTCOME1}^{adj} = \frac{Number\ of\ replicates\ where\ \min\ \{p_{OUTCOME1}^{rep}, p_{OUTCOME2}^{rep}\} < p_{OUTCOME1}^{raw}}{R}$$

$$p_{OUTCOME2}^{adj} = \max\ \left\{p_{OUTCOME1}^{adj}, \frac{Number\ of\ replicates\ where\ p_{OUTCOME2}^{rep} < p_{OUTCOME2}^{raw}}{R}\right\}$$

where $p_{OUTCOME}^{rep}$ is the *p*-value for an outcome in a particular replicate.

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

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

Exhibit A-1 shows the effect of this adjustment for the confirmatory outcomes reported in Exhibit 3-1. The first three columns of Exhibit A-1 are identical to those in Exhibit 3-1. The fourth column shows the unadjusted *p*-value without the multiple comparisons adjustment. The fifth column shows the *p*-value after we implement the adjustments described above. Consistent with the theory described earlier, the multiple comparisons adjustment increases the *p*-value for both estimates. The earnings impact estimate is insignificant prior to and after the adjustment. The SSDI benefits paid impact estimate moves from providing confirmatory evidence prior to the adjustment to providing marginal evidence after the adjustment (that is, the *p*-value moves from being statically significant at the 5 percent level to being statistically significant only at the 10 percent level after the adjustment).

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

	T1 Mean (1)	C1 Mean (2)	Impact Estimate (3)	*p*-value (Unadjusted) (4)	*p*-value (Multiple Comparisons Adjustment) (5)
Earnings Outcomes (January–December 2011)					
Total earnings **(confirmatory)**	$1,195	$1,204	-$9 ($25)	0.730	0.746
Total SSDI benefits paid **(confirmatory)**	$7,531	$7,508	$23* ($10)	0.040	0.082

Source: Analysis of SSA administrative records from the MEF, BODS, MBR, 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. See Chapter 3 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. The unadjusted *p*-value in column 4 shows the statistical test prior to the multiple comparison adjustment. The adjusted *p*-value in column 5 shows the statistical test after the multiple comparison adjustment.

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

A.3. Covariates

Exhibit A-2 lists the covariates included in the estimation of Equation (2) in Section A.1.

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 (Average Indexed Monthly Earnings) as of May 2011 (squared)
AIME (Average Indexed Monthly Earnings) as of May 2011 are equal to zero
Any employment in 2010 (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 RA 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
Monthly benefit amount (MBA) at baseline is equal to zero
Number of years receiving SSDI
Number of years receiving SSDI (squared)
Primary impairment category: Neoplasms Mental disorders Back or other musculoskeletal Nervous system disorders Circulatory system disorders Genitourinary system disorders Injuries Respiratory Severe visual impairments Digestive system Other impairments Unknown impairments
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.4. Sample Adjustments and Analysis Weights

This section describes the adjustments to the Stage 1 sample and the construction of the analysis weights used for calculating descriptive statistics and impact estimates. We use analysis weights in the estimation of program impacts in order to produce estimates for the national population of SSDI beneficiaries. These weights take account of the differing probabilities of selection into the sample for the different study sites and beneficiary subpopulations. Our final analysis weight also incorporates a contamination adjustment. Below, we describe the basic construction of the weight and the final adjustment made for contamination.

A.4.1. Adjustments to Analysis Sample

As shown in Exhibit A-3, our team made two adjustments to the original evaluation sample, one to account for deaths prior to random assignment, and one because of potential "contamination" because beneficiary pairs on the same primary record were assigned to different random assignment groups. As shown in column 1, random assignment yielded 79,991 T1 subjects, 79,991 C1-core subjects, and a large remaining pool of supplemental C1 subjects (827,817). In column 2, we show the adjustment for the sample to account for deaths. Specifically, SSA sent an update to the BOND sample in April 2012 that allowed our team to retrospectively identify T1 and C1 subjects who never were in BOND because they had died as of May 1, 2011 (one day prior to random assignment). These cases accounted for less than 1 percent of the overall sample. After this adjustment, the Stage 1 evaluation sample included a total of 822,331 subjects, spread across T1 (79,440 subjects) and C1 (901,709 subjects). This sample was used in the *Stage 1 Early Assessment Report*. Finally, in column 3, we show the contamination adjustment to the evaluation sample in column 2. The contamination is tied to the presence of BOND subjects who are on the same beneficiary records for eligibility but are in different random assignment groups. Specifically, the related subjects may influence the behavior of other subjects through example, through persuasion, or through program rules that directly tie the benefits of some BOND subjects together.[37] We dropped the contaminated BOND subjects, which affected less than 4 percent of BOND subjects. This approach is most consistent with a national offset policy, whereby no family would have different rules for different family members who receive SSDI. Given the large size of the C1 group relative to the T1 group, it is important to note that the probability that a subject is a member of a contaminated family varies by the size of the random assignment group; the probability of having a contaminated family member is higher in the T1 group relative to the C1 groups (core and supplement). This is most evident from the fact that more T1 subjects than C1-core subjects are dropped due to contamination (2,876 versus 1,387), even though the size of the T1 and C1-core groups are roughly the same (see Exhibit A-3). We adjusted the

[37] Under SSA rules, the earnings of the parent can affect the benefit level of the DAC, which has important implications if T1 and C1 subjects have related records. For example, a T1 primary beneficiary could increase his or her earnings in response to the benefit offset, which could influence both the primary and other auxiliary beneficiary's benefits, including a C1 DAC. If the parent's earnings change in response to the offset and in turn alter the DAC's benefit, the DAC's behavior might also change. If this happened, the DAC would be a "contaminated" control subject, because the DAC's circumstances would be affected by the BOND intervention. Another avenue for contamination under this same random assignment scenario is that the parent might factor in how his or her earnings would affect the benefits of the DAC. To fully understand how the DAC's benefits would be affected, the parent would need to consider the standard benefit rules for C1 subjects. This would result in the parent being a contaminated treatment subject, who is supposed to be making decisions in a program in which the offset exists for everyone. The same two avenues would have the potential for contamination if the assignments of the DAC and the parent were reversed.

weights for contamination to account for the differential probability of contamination by group, thereby ensuring that the results represent the full SSDI population.

For the purposes of this adjustment, we defined a family as two or more beneficiaries entitled to SSDI benefits on the basis of the work history of a common primary beneficiary and served by the same SSA area office. The most common example is a primary worker beneficiary (the parent) coupled with a DAC on the primary beneficiary's record. Another example is that of sibling DACs, identified because their benefits are based, at least partly, on the eligibility of a common primary beneficiary—a parent who receives Social Security disability or retirement benefits, or who is deceased.

Almost all of the families identified were pairs. We retained family pairs in the sample if both beneficiaries were randomly assigned to the same demonstration group. We dropped both of the beneficiaries from the sample if they were assigned to different groups. Pairs that were retained in the sample were weighted to reflect the probability of both beneficiaries being assigned to the same group. In essence, these weights allow the retained pairs to represent the "contaminated" pairs that were dropped from the analysis. Therefore, the BOND impact results extend to family clusters of two related BOND-eligible beneficiaries who are served by the same SSA area office.

In addition to the "contaminated" pairs, families with three BOND-eligible members or more were excluded from the analysis. The probability of all family members being assigned to the T1 group was so low that after "contaminated" families were removed from the sample, there were not enough of these larger families left to analyze (in fact, only a single family of three members remained in T1). This single family of three represents about 1 percent of beneficiaries in these larger families originally assigned to T1. In contrast, about 72 percent of the beneficiaries in these larger families remained in C1 after "contaminated" families were removed from the sample. Given this discrepancy, and the very large weights it would have implied for the three T1 subjects, all of these larger families from T1 and C1 were removed from the analysis sample. Beneficiaries from families with three or more BOND-eligible members represent a very small portion of all SSDI beneficiaries (about 0.5 percent of all prospective BOND subjects are in families of three or more BOND-eligible members). Their exclusion from the sample implies that BOND impact results do not generalize to the approximately 0.5 percent of SSDI beneficiaries who are in families of three or more beneficiaries served by the same SSA area office.

As will be described below, we generated separate weights for columns 2 and 3 in Exhibit A-3, in order to test the sensitivity of our findings to the contamination adjustment. The contamination-adjusted weight uses the same weight in column 2, except it adjusts weights on the beneficiary pairs that were retained to reflect the joint probability of both being assigned to the same group (i.e., the probability of being retained in the analysis sample).

Exhibit A-3. Stage 1 Evaluation Analysis Sample

	Initial Random Assignment Sample (1)	Analysis Sample after Adjustment for Mortality (2)	Final Analysis Sample (Adjusted for Mortality and Contamination) (3)	Cases Dropped (4)
T1	79,991	79,440	77,115	2,876
C1	907,808	901,709	891,598	16,210
C1-core	79,991	79,378	78,604	1,387
C1-supplement	827,817	822,331	812,994	14,823

Source: BOND Operations Data System (BODS).

Notes: Unless otherwise noted, all impact estimates in this report are based on the sample shown in Column 3. In the Appendix, we test the sensitivity of the impact findings to the use of the C1-core group and the inclusion of the sample in Column 2. The population size represents the national beneficiary population in the month of random assignment, which is the same for T1s and C1s (6,502,029 beneficiaries)

A.4.2. Construction of Analysis Weights

The first component of the analysis weight is the reciprocal of the probability of site selection. As explained in Stapleton et al. (2010), 10 SSA area offices were selected as sites for BOND from eight strata defined by census region (Northeast, Midwest, South, or West) and proportion of beneficiaries living in Medicaid buy-in states (low or high). A single area office was selected from each stratum, with one exception; two area offices were selected from the low Medicaid Buy-in stratum in the South region, which had many more area offices and beneficiaries than the other strata. [38] The area offices were selected in each stratum using probability proportional to size systematic sampling, in which size is defined as the number of SSDI beneficiaries served by the area office.

The second component of the analysis weights is the reciprocal of the probability of selection into T1 or C1 assignment groups. Within BOND sites, random assignment of beneficiaries into these groups occurred within six strata based on distinctions of short-duration beneficiaries (36 months or fewer) versus longer-duration beneficiaries (37 months or more), SSDI-only beneficiaries versus concurrent beneficiaries, and (for SSDI-only beneficiaries) Stage 2-eligible versus Stage 2-ineligible.[39] Thus, the six strata are:

- Short-duration SSDI-only who were Stage 2-eligible
- Short-duration SSDI-only who were not Stage 2-eligible

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

[39] All concurrent beneficiaries were ineligible for Stage 2. SSDI-only beneficiaries were ineligible for Stage 2 if they did not reside within BOND site areas, they resided in the Upper Peninsula of Michigan (a remote corner of the Wisconsin site where it was not practical to deliver EWIC services), or they had a legal guardian who was not an individual representative payee.

- Short-duration concurrent
- Long-duration SSDI-only who were Stage 2-eligible
- Long-duration SSDI-only who were not Stage 2-eligible
- Long-duration concurrent

For the T1 group, short-duration beneficiaries were oversampled such that one-half of the total T1 group is short-duration beneficiaries. The relative proportions of SSDI-only and concurrent beneficiaries in the T1 group are at their naturally occurring proportions within the BOND sites. The much larger C1 group includes at least as many beneficiaries in each of these strata as T1 but has relatively more long-duration beneficiaries and relatively more concurrent beneficiaries than T1.[40]

Below, we specify weights separately for (1) Stage 1 subjects who are unrelated to other prospective BOND subjects and (2) Stage 1 subjects who are related to another subject in the same assignment group. Each Stage 1 sample member who is unrelated to other prospective BOND subjects is assigned an analysis weight given by:

$$w_{mkjgi} = \left(\frac{N_m}{N_{mk}}\right)\left(\frac{N_{mkj}}{N_{mkjg}}\right)$$

where:

- w_{mkjgi} is the Stage 1 analysis weight for a beneficiary who is served by site *k* within national stratum *m*, is a beneficiary of type *j*, and has been randomly assigned to group *g*,
- N_m denotes the number of SSDI beneficiaries in stratum *m*,
- N_{mk} denotes the number of SSDI beneficiaries served by site *k* within stratum *m*,
- N_{mkj} denotes the number of SSDI beneficiaries served by site *k* within stratum *m* who are from one of the six possible strata defined above,
- N_{mkjg} denotes the number of SSDI beneficiaries of type *j* in site *k* within stratum *m* who are assigned to group *g* (T1 or C1).

In essence, the above expression is the product of a site weight and a within-site weight. Using this terminology, we can define the analysis weight of Stage 1 sample members who are related to another

[40] The T1 and C1-core groups were randomized on a one to one basis; hence, they include the same relative proportion of beneficiaries in each stratum. The much larger C1 group, which includes the C1 supplement subjects who were not included in the Stage 2 solicitation pool, has 1) relatively more concurrent beneficiaries than T1 because concurrent beneficiaries were not eligible for Stage 2 and 2) relatively more long-duration beneficiaries because of the oversampling of short-duration beneficiaries for T1 and the Solicitation Pool.

subject in the same assignment group as the product of the common site weight and the within site weights of each of the related sample members. In notation, this is:

$$w_{mkjgi} = \left(\frac{N_m}{N_{mk}}\right)\left(\frac{N_{mkj^i}}{N_{mkj^i g}}\right)\left(\frac{N_{mkj^r}}{N_{mkj^r g}}\right)$$

where:

- w_{mkjgi} , N_m , and N_{mk} are defined as above,
- N_{mkj^i} is equivalent to N_{mkj} defined above, with superscript *i* added to the type *j* to emphasize that this is the type *j* of beneficiary *i*,
- $N_{mkj^i g}$ is equivalent to N_{mkjg} defined above, with superscript *i* added to the type *j* to emphasize that this is the type *j* of beneficiary *i*,
- N_{mkj^r} denotes the number of SSDI beneficiaries served by site *k* within stratum *m* who are of the type *j* of beneficiary *r*, who is the related family member of beneficiary *i*,
- $N_{mkj^r g}$ denotes the number of SSDI beneficiaries served by site *k* within stratum *m* who are of the type *j* of beneficiary *r* (related family member of beneficiary *i*) who are assigned to group *g* (T1 or C1).

Note that related family members (beneficiary *i* and beneficiary *r*) who remain in the sample always are from the same stratum *m*, site *k*, and group *g* (otherwise they have been removed from the analysis sample). The related family members may differ only according to type *j*.

A separate set of analysis weights was created for the T1 versus C1-core impact analysis. For T1 subjects, the weights were identical to those described above. For C1 subjects, the related beneficiary pairs were considered contaminated if both members were not assigned to the C1-core. The weights for C1-core subjects were defined in a manner analogous to that above, with the definition of *g* being changed to T1 or C1-core (rather than T1 or C1).

A.5. Sensitivity Tests for Findings in Exhibit 3-1

Exhibit A-4 presents impact estimates for all beneficiaries when no BOND-eligible family members are excluded from the sample. The most notable change is that the estimated impact on the mean SSDI benefit paid is now $9 and statistically insignificant, compared to a marginally significant $23 in Exhibit 3-1. Additionally, the estimated impact on months with SSDI benefits paid is negative (-0.02 months over the eight-month period) and very significant, compared to an insignificant 0.00 in Exhibit 3-1. The sign of this estimate is opposite of the sign expected if the impact on mean SSDI benefits paid is positive. Finally, the estimate of the mean impact on SSI benefits is now a marginally significant -$6, compared to an insignificant -$2 in Exhibit 3.1. Although there are some changes in signs and significance for the estimates, all of these changes appear to be immaterial from a substantive perspective.

We also produced estimates using only C1-core subjects and compared them to estimates using the full C1 sample in order to verify that the weights developed for the latter were appropriately adjusting for that

sample's complex selection methodology (Exhibit A-5). Each point estimate changes by just a very small amount (compare the first two columns), as expected. Also as expected, the standard errors are substantially larger when only the C1-core subjects are used.

Exhibit A-4. Stage 1 Impact Estimates on Earnings and Benefit Outcomes Including All C1s Subjects, Including Contaminated Subjects

	T1 Mean	C1 Mean	Impact Estimate	Estimate from Exhibit 3-1
Earnings Outcomes (January–December 2011)[a]				
Total earnings **(confirmatory)**	$1,183	$1,198	$-14 ($19)	-$9 ($25)
Employment during year	16.14%	15.96%	0.18 (0.10)	0.13 (0.10)
Earnings above BYA	2.44%	2.40%	0.04 (0.10)	0.02 (0.12)
Earnings above 2 x BYA	0.94%	0.0.97%	-0.03 (0.05)	-0.03 (0.05)
Earnings above 3 x BYA	0.52%	0.52%	-0.01 (0.19)	0.00 (0.03)
Benefit Outcomes (May–December 2011)				
Total SSDI benefits paid **(confirmatory)**	$7,500	$7,491	$9 ($9)	$23* ($10)
Number of months with SSDI payments	7.47	7.48	-0.02*** (<0.01)	0.00 (<0.01)
Total SSI benefits paid	$338	$344	$-6* ($3)	-$2 ($5)
Number of months with SSI payments	1.37	1.38	-0.01 (<0.01)	-0.00 (<0.01)

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

Notes: All statistics are for the weighted analysis samples without an adjustment for contamination. Standard errors are in parentheses. Unweighted sample sizes: T1 = 79,440; C1 = 901,709. See Chapter 3 for variable definitions. Impact estimates are regression-adjusted. Benefit impacts are for the period from the date of random assignment (May 1, 2011) through December 2011, whereas employment and earnings impacts are for the full calendar year. Total earnings and SSDI benefits paid are the two confirmatory impacts, and statistical tests for the impacts on these two outcomes used multiple comparison adjustments. 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/.05/.01 levels, respectively, using a two-tailed t-test.

Exhibit A-5. Stage 1 Impact Estimates on Earnings and Benefit Outcomes Using C1-Core as a Comparison Group

	T1 Mean	C1-Core Mean	Impact Estimate	Estimate from Exhibit 3-1
Earnings Outcomes (January–December 2011)[a]				
Total earnings**(confirmatory)**	$1,195	$1,211	-$16 ($34)	-$9 ($25)
Employment during year	16.15%	16.07%	0.09 (1.43)	0.13 (0.10)
Earnings above BYA	2.43%	2.39%	0.04 (0.16)	0.02 (0.12)
Earnings above 2 x BYA	0.95%	0.98%	-0.03 (0.06)	-0.03 (0.05)
Earnings above 3 x BYA	0.53%	0.52%	0.01 (0.04)	0.00 (0.03)
Benefit Outcomes (May–December 2011)				
Total SSDI benefits paid **(confirmatory)**	$7,531	$7,505	$26 ($14)	$23* ($10)
Number of months with SSDI payments	7.49	7.51	-0.01* (0.01)	0.00 (<0.01)
Total SSI benefits paid	$340	$339	$1 ($6)	-$2 ($5)
Number of months with SSI payments	1.37	1.38	-0.01 (0.01)	-0.00 (<0.01)

Source: Analysis of SSA administrative records from the MEF, BODS, MBR, 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 = 78,604. See Chapter 3 for variable definitions. Impact estimates are regression-adjusted. Benefit impacts are for the period from the date of random assignment (May 1, 2011) through December 2011, whereas employment and earnings impacts are for the full calendar year. Total earnings and SSDI benefits paid are the two confirmatory impacts, and statistical tests for the impacts on these two outcomes used multiple comparison adjustments. 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/.05/.01 levels, respectively, using a two-tailed t-test.