BOND Implementation and Evaluation

Insights on Racial and Ethnic Equity

Deliverable 24f.5

Revised December 8, 2022

Submitted To:
Social Security Administration
Attn: Ms. Monisha Martinez Pardo
Office of Research, Demonstration, and Employment Support
6401 Security Boulevard
Baltimore, MD 21235

Contract No. SS00-10-60011

Prepared by
Hassan Enayati
Judy Geyer
Michelle Wood
Amanda Grittner

Report Context

SSA's Benefit Offset National Demonstration (BOND) examined the impacts of alternative SSDI program rules governing work. Specifically, BOND tested a benefit offset policy. BOND included two stages: a large nationally representative sample of SSDI beneficiaries in Stage 1 and a smaller sample of volunteers in Stage 2. Previous BOND reports presented findings from the process, participation, impact, and cost benefit analyses.

In its Equity Action Plan, the Social Security Administration (SSA) laid out its strategic approach to addressing equity in its programs and policies. The plan features data collection and analysis related to equity as a core objective. At the current time, SSA faces data limitations in that it does not have race and ethnicity information on a large proportion of its claimants and beneficiaries of the Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) programs, particularly for younger individuals.

The SSDI and SSI programs serve marginalized populations who face a multitude of systemic barriers. In a review of lessons learned from SSA demonstrations, Nichols and Hemmeter (2021) emphasized that none of the reviewed demonstrations assessed whether race or ethnicity was related to enrollment or delivery of services. Filling our knowledge gap on racial and ethnic variations within demonstrations supports the objectives of SSA's Equity Action Plan as well as President's Executive Order on Advancing Racial Equity and Support for Underserved Communities through the Federal Government.

This report uses existing data from BOND to examine racial and ethnic variations of a subset of demonstration participants and non-participants. Specifically, the analyses focus on the Stage 2 sample, which represents SSDI beneficiaries likely to volunteer for a benefit offset intervention. It considers three questions: (i) to what extent is there intersectional variation within the sample; (ii) to what extent does the impact of the benefit offset vary by race and ethnicity; and (iii) to what extent can we leverage existing data to predict race and ethnicity of eligible participants who elected not to volunteer in the demonstration?

Intersectional Analysis

Our exploration of the BOND Stage 2 volunteers examines variations in demographic characteristics, education, employment history, and work limitations by the intersection of race, ethnicity, and sex. After conducting a wide range of descriptive analyses, we identify five themes: (i) evidence of intersectional variations in a few demographic dimensions including age and marital status; (ii) clear patterns of a positive female education gap; (iii) wide employment variations in terms of current employment, number of hours worked, and use of workplace accommodations; (iv) Black and White men's positive self-reported health differentials drive the broader male positive differential; and (v) stark variations in perceptions of factors that limit ability to work including caring for others and transportation.

Differential Impacts on SSDI Benefits Due and Offset Use by Race and Ethnicity

Building on the robust impact analyses of the BOND benefit offset intervention reported in Gubits et al., 2018, we conduct exploratory analyses of the differential impacts of the benefit offset on SSDI benefits due and offset use between pairwise comparisons of Black, Hispanic, and White Stage 2 volunteers. Only the Black-White differential revealed statistically significant estimates. We find that the benefit offset had larger average impacts for Black subjects than for White subjects on the amount of benefits due and the percent of subjects with any benefits due. We attribute this pattern to differences, at baseline, in average age and employment that vary by race.

Predicting Race and Ethnicity

The analysis to predict race and ethnicity aims to address two goals. First, we will examine the feasibility and accuracy of an algorithm that predicts race and ethnicity using SSA data. Second, we apply those predictions to explore the differences between the Stage 2 volunteers and non-volunteers. We use name and geocode data along with the Modified Bayesian Improved First Name, Surname, and Geocoding (BIFSG) methodology to predict race and ethnicity of the Stage 2 volunteers and non-volunteers. We make five key observations from this analysis: (i) this algorithm can accurately predict race and ethnicity among SSDI beneficiaries and there is compelling evidence that additional calibration of the estimates using self-reported race and ethnicity data from a small proportion of the sample improves the accuracy of the prediction; (ii) Stage 2 volunteers are more likely to be Black and less likely to be White than the non-volunteers in the solicitation pool; (iii) the higher proportions of Black women to Black men as well as the lower proportions of Hispanic and White women to Hispanic and White men among Stage 2 volunteers seen in the intersectional analysis holds for the Stage 2 non-volunteers; (iv) there is evidence of variation in short duration status and age among racial and ethnic groups, even for the Stage 2 non-volunteers; and (v) even after accounting for a set of individual characteristics, select BIFSG estimates have a statistical relationship with sex and the decision to volunteer.

Table of Contents

List of Exhibits

Acronyms Used in This Report

API	Asian Pacific Islander
AIAN	American Indian/Alaska Native
BIFSG	Bayesian Improved First Name, Surname, and Geocoding
BOND	Benefit Offset National Demonstration
C1	Stage 1 control group subjects (subject to current law)
C2	Stage 2 control group subjects (subject to current law)
EAP	Equity Action Plan (developed by the Social Security Administration)
EWIC	Enhanced work incentives counseling
MBR	Master Beneficiary Record
RECS	Race Ethnicity Collection System
SSA	Social Security Administration
SSDI	Social Security Disability Insurance
SSI	Supplemental Security Income
T1	Stage 1 treatment subjects (subject to the offset rules and offered WIC)
T21	Stage 2 treatment subjects (subject to the offset rules and offered WIC)
T22	Stage 2 treatment subjects (subject to the offset rules and offered EWIC)
WIC	Work incentives counseling

1. Introduction

1.1. Motivation

In response to Executive Order 13985[1] the Social Security Administration is examining its programs and policies to advance equity. The agency has developed an Equity Action Plan (EAP) to guide these efforts. As SSA leaders noted in a June 16, 2022, Equity Action Plan Stakeholder Engagement Meeting (SSA 2022), one focus area for the EAP is "Improving Data Collection and Conducting Analyses to Identify Inequities in Programs." The EAP emphasizes that collecting and analyzing information on racial and ethnic diversity among program applicants and beneficiaries is essential if SSA is to assess and promote equity in its programs and services.

While SSA aims to make progress on this priority, a substantial amount of missing data on race and ethnicity, especially for younger individuals, limits what SSA can learn about diversity among its program applicants, beneficiaries, and demonstration participants. SSA stopped publishing data on the racial composition in the SSI program after 2002 and in the Old-Age, Survivors, and Disability Insurance program after 2009 (Martin and Murphy 2014; Martin 2016). SSA took this step largely because of changes to the process for assigning new Social Security numbers, called enumeration. Nearly all individuals (96 percent) now receive their original Social Security Numbers as part of the hospital-birth registration process called Enumeration at Birth. States administer this process, and SSA does not receive information on race and ethnicity. In the 2022 EAP, SSA estimates that it currently has the race and ethnicity for approximately 59 percent of living Social Security Number holders.[2]

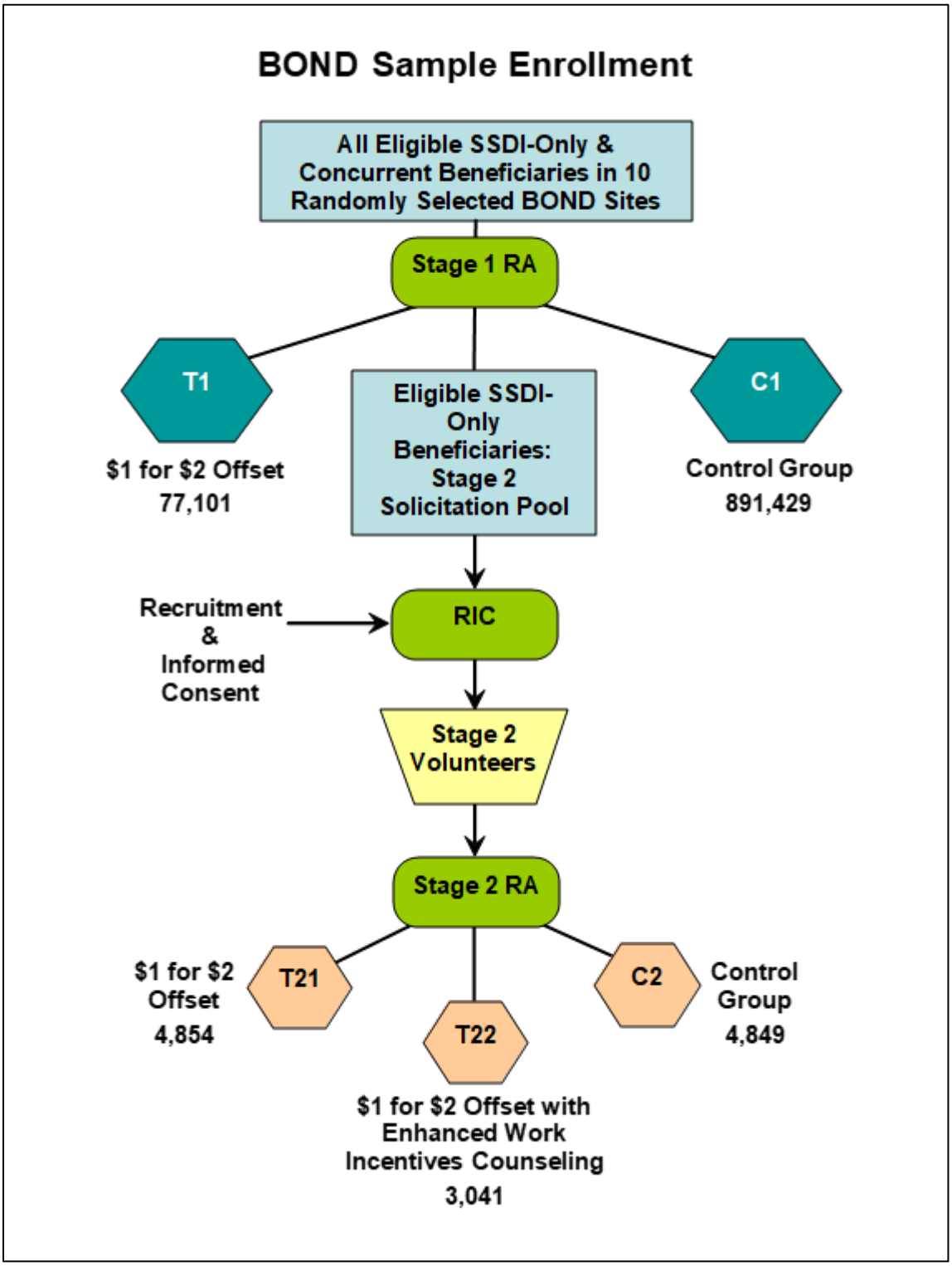
Data from SSA's recent Benefit Offset National Demonstration (BOND) provide a unique opportunity for SSA to consider the racial and ethnic equity within its programs. BOND tested changes to Social Security Disability Insurance (SSDI) program rules governing work and other supports. BOND incorporated a $1 for $2 benefit offset allowing beneficiaries to retain some of their monthly cash benefit while working. Exhibit 1-1 illustrates the underlying beneficiary population eligible to participate in BOND and the assignment process to demonstration groups.

---

[1] Executive Order 13985, "Advancing Racial Equity and Support for Underserved Communities Through the Federal Government," Jan. 20, 2021, https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/.

[2] For people born before 1987, SSA collected race and ethnicity data from persons who voluntarily provided it when applying for an original or replacement SSN card. This voluntarily-provided race and ethnicity data is stored in SSA's Race Ethnicity Collection System (RECS). Since 1987, SSA has issued most original SSN cards through the enumeration at birth process. Qualified immigrants receive most of the remaining new SSNs upon entry into the United States. The Department of Homeland Security administers this procedure, called Enumeration at Entry, which does not collect information on race and ethnicity. SSA's Equity Action Plan notes that as a result of current enumeration processes, SSA now has race and ethnicity data for only about 59 percent of living SSN holders. In November of 2021, SSA issued guidance to its frontline staff to encourage individuals who apply for new or replacement SSN cards in an SSA field to provide race and ethnicity information. The Equity Action Plan can be accessed at https://www.ssa.gov/open/materials/SSA-EO-13985-Equity-Action-Plan.pdf.

Exhibit 1-1. Overview of BOND Sample and Random Assignment Process



Source: Exhibit modified from Gubits et al. (2018).
Note: DI = disability insurance; RA = random assignment; RIC = recruitment and informed consent.

The two stages of BOND are summarized below.

- **Stage 1** tested how a national benefit offset would affect earnings and program outcomes for the entire SSDI population. The Stage 1 sample was a nationally representative cross-section of the SSDI population under age 60 as of May 2011 in 10 randomly selected SSA area offices (Stapleton et al. 2010).[3] At each BOND site, all current SSDI beneficiaries between ages 20 and 59 receiving benefits based on disability and who were not part of another SSA demonstration were included in the BOND sample. In Stage 1, the demonstration randomly assigned beneficiaries into a treatment group subject to benefit offset rules and offered regular work incentives counseling; a current-law control group; or to a Stage 2 solicitation pool (i.e., the potential SSDI-only participants) that received outreach and recruitment to volunteer for Stage 2.

- **Stage 2** tested the impact of the $1 for $2 benefit offset for those expected to be most likely to use the offset—recruited and informed volunteers. Stage 2 also tested the extent to which enhanced counseling affected impacts. In Stage 2, the demonstration randomly assigned volunteers into one of three assignment groups: a treatment group subject to the benefit offset rules and offered regular work incentives counseling, a second treatment group subject to the benefit offset rules and offered enhanced work incentives counseling; or a current-law control group.

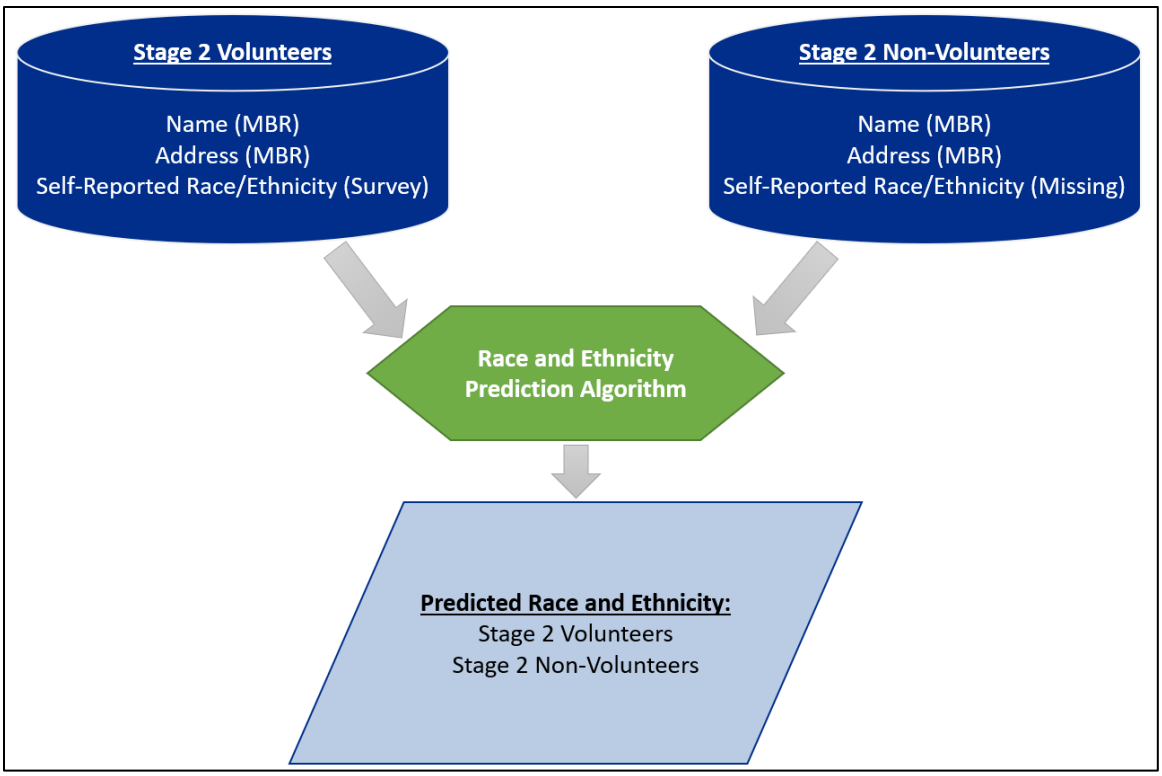
Several features of the BOND data motivate this equity-focused report. Analysis of the Stage 2 sample offers SSA insights on a group of SSDI beneficiaries eager to participate in a benefit offset program. Along with detailed information about previous work history, health, and other topics, SSA collected self-reported race and ethnicity data from 12,869 SSDI beneficiaries who volunteered for Stage 2 of BOND (Gubits et al. 2013). While the 10 BOND sites collectively account for a random sample of approximately 20 percent of the SSDI population in 2011, the SSDI beneficiaries in these locations are not necessarily representative of the national SSDI population in demographic characteristics such as age, sex, race, or ethnicity. Thus, the BOND Stage 1 sample is representative in statistical expectation, but it does not necessarily reflect the actual racial and ethnic composition of the SSDI population. This nuance is another example of why it is useful for SSA to think about who participates in each demonstration it conducts. Because the composition of participants in any given demonstration will vary based on location, eligibility, and other factors, an equity perspective requires an understanding of the variation between the characteristics of the eligible population for the demonstration and the enrolled participants. The analysis in this report highlights the racial composition within one specific group of volunteers in the BOND demonstration. While not generalizable to Stage 1 participants or to other groups of volunteers for SSA demonstrations, this analysis showcases the value of conducting racial analyses.

Although the self-selected Stage 2 volunteers do not represent the general SSDI caseload, descriptive analysis of the variation in demographic characteristics within this group can build SSA's understanding of the racial and ethnic diversity within the group of BOND volunteers. Previous analyses conducted through the BOND evaluation have not described the racial and ethnic composition in the group of Stage 2 volunteers in detail. A deeper understanding of the racial and ethnic composition of a group of SSDI

[3] The Abt evaluation contractor team randomly selected the 10 BOND sites from 53 SSA Area Offices. The 10 sites covered seven full states (Alabama, Arizona, Colorado, Maine, New Hampshire, Vermont, and Wyoming) plus the District of Columbia. The sites also included substantial portions of nine additional states (California, Florida, Maryland, Massachusetts, Michigan, New York, Texas, Wisconsin, and Virginia) and smaller portions of two other states (Pennsylvania and West Virginia).

beneficiaries interested in a benefit offset work incentive may offer insights on how these types of interventions intersect with structural barriers tied to race, ethnicity, and sex of those most interested in using those incentives. In addition, the race and ethnicity data collected from the Stage 2 volunteers offer opportunities to test new procedures for predicting race and ethnicity that could hold promise for SSA's equity work. Exhibit 1-2 illustrates the sample, data, and broad approach used to predict race and ethnicity in this report. The evidence we present on the prediction of race and ethnicity can supplement the other analyses SSA is conducting as described by the Acting Associate Commissioner of the Office of Research, Evaluation, and Statistics, Dr. Natalie Lu, at the June 2022 SSA EAP Stakeholder Engagement Meeting.

Exhibit 1-2. Overview of Race and Ethnicity Prediction Approach



Note: MBR = Master Beneficiary Record; Survey = Stage 2 Baseline Survey; Missing refers to the lack of race and ethnicity data within the MBR.

1.2. Goal

To support SSA's priority under the EAP to explore equity in its programs and services, this report draws on the Stage 2 baseline survey data to provide a more detailed description of the BOND Stage 2 sample than previously available. The report also offers insights that might inform other efforts at SSA to determine whether its demonstrations and services are reaching a diverse group of beneficiaries. To support this goal, we examine three research questions. The first two questions aim to better understand features of BOND. The third question leverages BOND data to explore a new methodology that SSA can use to predict race and ethnicity.

Examination of BOND Participants and Differential Impacts of the Benefit Offset

1. To what extent do demographic characteristics, education, employment history, and work limitations vary among self-reported intersectional groups of race, ethnicity, and sex within the Stage 2 volunteers?

2. To what extent does the impact of the BOND benefit offset on SSDI benefits due and offset use vary by self-reported race and ethnicity?

Examination of the Feasibility of a Race and Ethnicity Imputation Method for the SSDI Population

3. To what extent can we leverage existing data on Stage 2 volunteers to predict the race and ethnicity of Stage 2 non-volunteers in order to examine the racial and ethnic composition variations between the two groups?

To address the first research question, we examine the Stage 2 baseline survey data to look for variations between self-reported identity groups (i.e., the race, ethnic, and sex intersectional groups) in baseline characteristics such as employment history, self-reported health status, self-reported barriers to work, and understanding of SSDI rules. This analysis offers the opportunity to determine whether systemic variations exist between racial and ethnic groups within the group of Stage 2 volunteers. This analysis may contribute to SSA's efforts to scrutinize policies and practices as called for in the Equity Action Plan.

Next, we leverage the measures of self-reported race and ethnicity to test for differential treatment effects of the BOND benefit offset in Stage 2. We conclude with a summary and set of next steps for SSA to consider.

Addressing research question three requires information about the race and ethnicity of the Stage 2 solicitation pool. However, those who were recruited but did not enroll did not complete a baseline survey. This missing data problem presents an opportunity to test the viability of using a name and geocode based algorithm to predict race and ethnicity. We implement the Modified Bayesian Improved First Name, Surname, and Geocoding (BIFSG) methodology (Voicu 2018). If found viable, this method is something that SSA might consider applying more broadly to use existing data to examine racial and ethnic representation in its programs. Here, we assess the accuracy of the BIFSG algorithm against the self-reported race and ethnicity data in the Stage 2 baseline survey by exploring the differentials in summary statistics and conducting sensitivity analyses. Findings from these analyses clarifies under what circumstances the BIFSG approach appears viable for predicting race and ethnicity in the SSDI population.

1.3. Prior Analysis of the Stage 2 Sample

The BOND evaluation has examined several questions about the Stage 2 volunteers. As reported in Gubits et al. (2018),

> The Stage 2 outreach and recruitment was intended to produce a select sample of SSDI beneficiaries, distinct from the national SSDI caseload in their likelihood to use the benefit offset. Differences in the 2011 employment rates confirm that the Stage 2 sample is indeed distinct from the Stage 1 sample. Altogether, 36 percent of Stage 2 control subjects were working in 2011, compared with 14 percent of Stage 1 control subjects. [Gubits et al. (2018); pages 8 and 9].

In an early assessment of Stage 2, Gubits et al. (2013) used SSA administrative data to compare Stage 2 volunteers to non-volunteers in the solicitation pool. They found that women volunteered for the demonstration at a higher rate than men and volunteers tended to be younger than non-volunteers, with a mean age of 47.6 years vs. 49.2 years for non-volunteers. Volunteers were generally representative of the solicitation pool in terms of primary impairment, although SSDI beneficiaries with mental health disorders were slightly more likely than other beneficiaries to volunteer. Beneficiaries with short-duration SSDI receipt (defined as having received benefits for 36 months or less) volunteered at higher rates than those with longer SSDI receipt, resulting in a mean duration among volunteers of 53.4 months compared to 73.2 months among non-volunteers. While monthly SSDI benefit amounts were similar between volunteers and non-volunteers, disabled adult children were less likely to volunteer as were those who had a representative payee. With the possible exception of women's over-representation in the volunteer group, these findings are unsurprising. Younger and shorter-duration beneficiaries may be more interested in work than those who are older and have longer experience of benefits receipt. Similarly, beneficiaries who do not require a representative payee may also be more able to work.

Gubits et al. (2013) also compared baseline characteristics across the three Stage 2 assignment groups and found that random assignment produced three well-matched groups. The baseline equivalence analysis confirmed the internal validity of the Stage 2 impact analysis.

Gubits et al. (2013) also examined data from the Stage 2 baseline survey to describe the volunteers. That analysis tabulated racial and ethnic composition of all volunteers and all volunteers in each assignment group. The analysis found that among all volunteers, more than half (55.2 percent) identified as non-Hispanic white, just over one quarter (26.4 percent) identified as Non-Hispanic Black, and 9.1 percent identified as Hispanic. Nearly 8 percent identified as another racial group and less than one percent identified as American Indian or Alaskan native, Asian, and Native Hawaiian or other Pacific Islander.

However, previous descriptive analysis did not cross-tabulate responses to the baseline survey by racial and ethnic groups to examine whether baseline characteristics differ by racial and ethnic identify groups within the group of volunteers. This paper takes a deeper look at the baseline characteristics, cross tabulating by race to determine whether employment history, health, and other characteristics vary by race and ethnicity.

2. Variations by Identity Groups

2.1. Intersectional Approach

The theoretical framework of intersectionality rests on the premise that individuals' lived experiences are influenced by their socio-demographic intersections, which themselves are shaped by social power structures (Collins 1991; Crenshaw 1989). Social scientists have been increasingly advocating for the application of intersectional methods in quantitative policy-focused research (Bauer and Scheim 2019; Whitebread, Dolamore, and Stern 2022). In this chapter, we apply a descriptive intersectional analysis framework to examine the extent to which demographic characteristics, education, employment history, and work limitations vary among self-reported intersectional groups of race, ethnicity, and sex within the Stage 2 volunteers. This analysis will not only provide a more complete picture of the Stage 2 sample than previously available but also provides insights for SSA as it considers larger questions about whether SSA's demonstrations and services are reaching a diverse group of beneficiaries.

We extend the prior analysis of the Stage 2 sample by considering intersectional variations in baseline characteristics and environmental factors by race, ethnicity, and sex. Our approach leverages descriptive analytics to better understand the similarities and dissimilarities of the Stage 2 sample. We compare baseline characteristics for the intersection of race-ethnicity and sex. We tabulate and report statistical differences for the following racial/ethnic groups: non-Hispanic Black, Hispanic with any race, and non-Hispanic White. While the data also include American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander as race categories, the sample sizes for these categories are too small to investigate variations across race and sex. In this report we will refer to non-Hispanic Black beneficiaries as Black beneficiaries and non-Hispanic White beneficiaries as White beneficiaries.

This set of analyses examines the sex differences within each of the three racial and ethnic groups studied in this report. Reported p-values inform us of the level of statistical significance for the observed difference within the racial group by sex. To contextualize the sex differences across racial groups, we will at times compare the sex-race/ethnicity estimates to sex-only or race/ethnicity-only estimates, which are reported in Appendix A.

2.2. Data Used in the Intersectional Approach

In this section, we analyze data from the Stage 2 baseline survey, collected for the Stage 2 volunteers. We also use SSA administrative data available for the Stage 2 solicitation pool. Stage 2 aims to examine the effect of the offset on the population most likely to use it; hence, its participants were recruited and informed volunteers. The solicitation pool for Stage 2 consisted of 238,070 SSDI-only beneficiaries. From this sample, Stage 2 randomly assigned 12,744 volunteers into one of three groups: offset plus regular work incentives counseling (WIC) T21 (4,849): Offset plus enhanced work incentives counseling (EWIC), T22 (4,854), and Stage 2 control group C2 (3,041).[4] The Abt implementation team recruited and enrolled Stage 2 volunteers from March 1, 2011 to September 28, 2012.

[4] Stage 2 volunteer analytic sample is based on the 12,954 randomly assigned volunteers. The Stage 2 volunteer analysis sample excludes 210 beneficiaries who are related to other BOND subjects to avoid contamination effects that might arise from the fact that almost all such beneficiaries (204 of the 210) were assigned to different BOND groups (see Appendix A of Gubits et al. (2018) for details).

During Stage 2 enrollment but before random assignment, the Abt evaluation team collected the BOND baseline survey data. The survey was designed to provide background characteristics of beneficiaries not available in the administrative records. In addition to race and ethnicity, it asked questions on demographics, family and housing status, employment, income, understanding of work incentives, and perception of the demonstration. We use these data to answer research question one. Of the 12,744 Stage 2 volunteers, we will focus on the 11,360 volunteers that responded as either Black, Hispanic, or White. Across the six intersectional groups considered in this analysis, there are 1,858 (16.4 percent) Black women, 1,435 (12.6 percent) Black men, 503 (4.4 percent) Hispanic women, 637 (5.6 percent) Hispanic men, 3,369 (29.7 percent) White women, and 3,558 (31.3 percent) White men.

2.3. Estimated Variations

We build upon earlier analyses of the Stage 2 sample, conducting further cross tabulations of the data. In this section, we highlight five themes that emerged:

- Evidence of intersectional variations in a few demographic dimensions including age and marital status;
- Clear patterns of a positive female education gap;
- Wide employment variations in terms of employment at recruitment[5], number of hours worked, and use of workplace accommodations;
- Black and White men's positive self-reported health differentials drive the broader male positive differential; and
- Large variations in perceptions of factors that limit ability to work including caring for others and transportation.

We start by examining the demographic variations across our six intersectional identity groups. Exhibit 2-1 shows that the male-female difference in the age distribution only holds for Black beneficiaries. Specifically, we find weak statistical evidence that Black men are more likely to be at the extremes of the distribution.[6] All men are more likely than women to be married with a 34.2, 28.9, and 6.5 percent differential for Black, Hispanic, and White men, respectively. Similar patterns hold for men with regard to living with their partner. Both Hispanic and Black men are more likely to speak a language other than English in the home than their female counterparts, though Hispanic participants as a group had the largest prevalence rate.

Women within each racial and ethnic group have higher educational outcomes than their male counterparts. For example, the Hispanic sex gap on having at least a high school degree favors women by 9.1 percentage points. The same gap for Black and White women is 3.5 percentage points, each. The differences in Bachelor's degrees or higher is significant and meaningful for Black (5.6 percentage

[5] Current employment refers to self-reported employment at the time the beneficiary completed the baseline survey.

[6] The Black male versus female averages are not significantly different from each other. That said, the broader age distributions, as measured in the specified age cohorts, are weakly statistically different, with a p-value of 0.055. Both the left and right tails are thicker for men than for women.

points) and White (13.1 percentage points) women. The sex gap finding that women are more likely to have been currently in school at baseline only holds for Black and White women.

Next, we apply our intersectional approach to consider ways in which the Stage 2 sample differed in their employment outcomes. Looking *across* all six groups, Black men have the lowest rate of current (measured at baseline) employment (16.4 percent) while White women have the highest (32.8 percent), as seen in Exhibit 2-2. Looking *within* racial groups, we find the rate of employment for Black women is 4.3 percentage points higher than the rate for Black men. White women have a 3.2 percentage point differential. This intersectional analysis highlights that the sex gap in hours worked and tenure by sex is driven by White men. Black women are the only group with a statistically significant within race difference in work accommodation. In fact, White women are nearly 2.5 times more likely than Black men to use an accommodation. Finally, the positive knowledge gap in knowledge of the program rules for women is observed in raw differences at the intersectional level but is mostly statistically insignificant.

Race by sex differentials extend into health outcomes. In Exhibit 2-3, we see the positive self-reported health gap for men holds across all races but is only significant for Black and White men. Hospital use has a noteworthy pattern. While Black men with overnight stays have shorter duration of hospital visits than Black women, the opposite is true for Hispanic and White women.

Our final set of observations focuses on self-reported barriers to work. Exhibit 2-3 shows that White women more often perceive a work limitation due to physical or mental condition than any other group and do so at a statistically significant difference from the second highest group, White men. There are sex gaps in limitations due to transportation for Black and White men, which corresponds to the evidence we see of lowest vehicle ownership rate for Black men. For each racial and ethnic group, the responsibility of caring for children or others is more likely to be a limiting factor to employment for women with percentage point differentials of 6.6, 4.5, and 4.1 for Black, Hispanic, and White women, respectively.

Exhibit 2-1. Baseline Demographic Characteristics by Self-Reported Race-Ethnicity and Sex

Characteristic	Black Women	Black Men	p-value	Hispanic Women	Hispanic Men	p-value	White Women	White Men	p-value
Number of Beneficiaries	1,858	1,435		503	637		3,369	3,558	
Age									
Age	46.97	47.23	0.613	46.68	47.09	0.521	47.69	47.42	0.356
Age category									
20-29	4.38	6.31	0.055 *	5.94	5.61	0.239	4.89	6.56	0.160
30-39	17.70	14.19		17.89	14.10		13.56	13.51	
40-44	13.35	12.14		12.99	14.21		12.38	11.24	
45-49	16.42	17.64		14.50	18.72		17.45	17.88	
50-54	23.64	23.43		24.34	22.53		24.98	23.32	
55-60	24.52	26.29		24.34	24.84		26.72	27.49	
Marital Status									
Married	18.68	25.06	0.000 ***	30.49	39.31	0.000 ***	33.23	35.38	0.000 ***
Widowed, Divorced, or Separated	45.33	35.86		47.08	32.42		41.71	31.01	
Never married	35.99	39.09		22.43	28.27		25.06	33.61	
Currently living with spouse or partner	22.05	32.79	0.004 ***	37.45	45.91	0.001 ***	39.45	42.03	0.381
Language Spoken at Home									
Primary language at home is not English	0.92	2.56	0.028 **	27.58	36.41	0.048 **	1.19	1.25	0.881
Educational Attainment									
High school degree or more	86.88	83.41	0.010 ***	82.68	73.61	0.007 ***	94.7	91.23	0.004 ***
College degree or more	30.41	20.79	0.000 ***	26.78	22.72	0.274	44.29	31.17	0.000 ***
Bachelor's degree or more	14.89	9.04	0.000 ***	11.59	11.98	0.881	25.06	16.84	0.000 ***
Currently In School									
Currently enrolled in school or taking classes	10.93	7.97	0.033 **	9.03	7.07	0.258	7.94	5.93	0.043 **
Currently working toward degree, certificate, or license	9.16	7.06	0.065 *	7.91	5.14	0.118	6.84	4.92	0.040 **
Full-time student	5.02	4.00	0.154	3.61	3.18	0.733	3.46	2.14	0.049 **
Part-time student	5.68	3.97	0.032 **	5.33	3.82	0.173	4.26	3.62	0.374

Source: SSA administrative records (MBR) and Stage 2 baseline survey.

Note: This analysis compares mean differences by sex within race/ethnicity. We test the hypothesis that the mean characteristic for women equals the mean characteristic for men within a specific race/ethnicity. If the characteristic is measured by a continuous or binary variable, we use a t-test to compare means and obtain p-values. If the characteristic is measured by a categorical variable, we use a Chi-Square test to compare the distributions and obtain p-values. Stars indicate statistical significance. One star indicates that the characteristics are statistically significantly different for men and women at a 10 percent significance level. Two stars indicate a significance level of 5 percent, and three stars indicates a significance level of 1 percent.

Exhibit 2-2. Baseline Employment-Related Characteristics by Self-Reported Race-Ethnicity and Sex

Characteristic	Black Women	Black Men	p-value	Hispanic Women	Hispanic Men	p-value	White Women	White Men	p-value
Number of Beneficiaries	1,858	1,435		503	637		3,369	3,558	
Labor Force Participation									
Currently working at a job	20.69	16.44	0.004 ***	22.21	19.45	0.315	32.77	29.58	0.043 **
Currently looking for work	28.25	30.96	0.314	26.95	27.82	0.871	30.61	31.98	0.566
Of those not working, # of months since last worked (Median)	40.09	39.02	0.836	38.66	39.37	0.833	42	39.49	0.986
Job characteristics (for those currently working)									
Hours worked per week (Median)	20.18	23.74	0.223	19.66	24.08	0.263	19.27	19.71	0.000 ***
Tenure (in months) at current job (Median)	12.09	12.31	0.129	9.71	18.7	0.211	15.24	18.34	0.002 ***
Annual Earnings									
$0	79.55	82.73	0.006 ***	76.39	81.15	0.094 *	67.67	70.81	0.020 **
$1-2,999	4.98	2.71		6.87	2.48		7.84	5.97	
$3,000-5,999	3.28	2.62		4.35	3.57		5.15	5.2	
$6,000-8,999	2.82	2.52		4.5	3.73		6.09	4.61	
$9,000-11,999	2.99	3.84		2.12	4.82		5.84	5.65	
$12,000-14,999	1.73	2.02		1.51	0.97		3.23	2.65	
$15,000 or above	4.65	3.56		4.26	3.29		4.19	5.12	
Work Accommodations									
Use of special equipment related to disability at work	5.52	3.37	0.002 ***	3.77	4.65	0.514	8.38	7.11	0.124
Use of personal assistance service at work	0.78	1.44	0.254	2.11	1.2	0.395	2.86	2.74	0.774
Ability to Work									
Had someone help with baseline interview	0	0.47	0.064 *	0.23	0.61	0.338	0.23	0.88	0.017 **
Knowledge of Program Rules									
Ever heard of trial work period (TWP)	72.64	70.7	0.333	57.07	57.03	0.990	79.72	77.73	0.075 *
Ever heard of extended period of eligibility (EPE)	20.85	17.08	0.016 **	15.43	14.47	0.528	25.1	23.98	0.417

Source: SSA administrative records (MBR) and Stage 2 baseline survey.

Note: This analysis compares mean differences by sex within race/ethnicity. We test the hypothesis that the mean characteristic for women equals the mean characteristic for men within a specific race/ethnicity. If the characteristic is measured by a continuous or binary variable, we use a t-test to compare means and obtain p-values. If the characteristic is measured by a categorical variable, we use a Chi-Square test to compare the distributions and obtain p-values. Stars indicate statistical significance. One star indicates that the characteristics are statistically significantly different for men and women at a 10 percent significance level. Two stars indicate a significance level of 5 percent, and three stars indicates a significance level of 1 percent.

Exhibit 2-3. Baseline Health- and Transportation-Related Characteristics by Self-Reported Race-Ethnicity and Sex

Characteristic	Black Women	Black Men	p-value	Hispanic Women	Hispanic Men	p-value	White Women	White Men	p-value
Number of Beneficiaries	1,858	1,435		503	637		3,369	3,558	
Self-reported Health Status									
Good, very good, or excellent	30.51	36.72	0.023 **	31.52	37.9	0.042 **	38.55	42.14	0.241
Hospital Use									
Stayed overnight in hospital in last 12 months	32.99	31.93	0.381	30.82	33.83	0.448	28.81	28.36	0.768
Number of nights in hospital in last 12 months (Median)	4.05	3.95	0.072 *	3.46	4.55	0.056 *	4.01	4.59	0.001 ***
Health Insurance									
Have health insurance	94.18	92.19	0.015 **	94.88	92.17	0.038 **	96.45	94.41	0.000 ***
Limitations (Agree or strongly agree responses)									
Inability to work because of a physical or mental condition	81.43	81.34	0.962	84.76	84.35	0.882	88.57	86.14	0.004 ***
Inability to work because I don't have reliable work transportation	19.51	25.94	0.000 ***	18.96	21.94	0.112	14.71	16.12	0.062 *
Inability to work because I am caring for children or others	14.41	7.78	0.001 ***	12.94	8.43	0.046 **	10.41	6.32	0.000 ***
Difficult to work because I am afraid I will lose disability benefits	35.69	39.27	0.269	45.25	48.36	0.567	42.02	41.21	0.649
Inability to work because I am finishing a school/training program	4.75	4.76	0.999	4.68	4.15	0.692	4.37	3.78	0.229
Workplaces are not accessible to people with my disability	46.12	45.96	0.953	52.36	47.35	0.217	42.66	40.7	0.230
Lacking skills or training I need to return to work	31.73	34.78	0.074 *	36.34	38.8	0.535	31.9	33.04	0.267
Difficulty to re-qualify for SS disability benefits in future if I work	38.98	39.18	0.938	45.12	41.78	0.643	39.59	39.06	0.689
Personal goals include moving up in a job or learning new skills	92.21	91.21	0.231	91.24	89.59	0.352	85.23	88.73	0.020 **
Usual Mode of Transportation									
Own car, truck, or van	60.04	52.56	0.001 ***	66.24	62.08	0.036 **	77.86	74.75	0.011 **
Public transportation	35.59	49.14	0.000 ***	27.83	36.19	0.004 ***	15.68	20.95	0.001 ***
Friends or relatives	64.15	56.61	0.004 ***	64.48	54.13	0.002 ***	49.88	44.08	0.000 ***
Walk	31.4	43.9	0.000 ***	31.78	37.26	0.123	26.15	31.52	0.001 ***
Taxi, van, or paratransit service	30.56	27.89	0.168	24.99	20.9	0.198	14.7	14.87	0.850
Wheel or motorized scooter	4.23	6.31	0.011 **	7.21	6.20	0.399	6.89	7.22	0.604
Other	1.00	5.65	0.001 ***	0.89	5.33	0.012 **	4.00	9.11	0.000 ***
Ability to Drive									
Able to drive a car	78.97	79.26	0.710	80.92	78.43	0.239	86.29	84.06	0.003 ***
Have a valid driver's license	77.38	68.02	0.000 ***	79.12	74.47	0.064 *	88.49	83.64	0.000 ***
Access to a car that runs	83.43	78.09	0.005 ***	86.37	87.29	0.777	91.64	91.93	0.756

Source: SSA administrative records (MBR) and Stage 2 baseline survey.

Note: This analysis compares mean differences by sex within race/ethnicity. We test the hypothesis that the mean characteristic for women equals the mean characteristic for men within a specific race/ethnicity. If the characteristic is measured by a continuous or binary variable, we use a t-test to compare means and obtain p-values. If the characteristic is measured by a categorical variable, we use a Chi-Square test to compare the distributions and obtain p-values. Stars indicate statistical significance. One star indicates that the characteristics are statistically significantly different for men and women at a 10 percent significance level. Two stars indicate a significance level of 5 percent, and three stars indicates a significance level of 1 percent.

3. Differential Effects of SSDI Benefits Due and Offset Usage

In this section, we shift focus from variations in the individual characteristics of Stage 2 volunteers across racial and ethnic groups to differential impact of the benefit offset tested in BOND for the Stage 2 sample. Mirroring the call in SSA's Equity Action Plan to examine equity from multiple perspectives, the prior section of this report asked questions about who participated in Stage 2 of BOND. The current section asks if the effect of the BOND intervention varied by self-reported race and ethnicity. The exploratory analyses in this chapter aims to shed light on any variation in impact by race and ethnicity though they are not held to the evidentiary standards of the full sample confirmatory outcomes reported in Gubits et al. 2018.

3.1. Prior Subgroup Impact Analysis

The main findings of the BOND benefit offset intervention for the Stage 2 sample identified no statistically significant evidence of an effect on average earnings but did reveal a four percent increase in the average amount of SSDI benefits due (Gubits et al. 2018). Part of this increase is due to the windfall of partial benefits to beneficiaries who would have otherwise received zero benefits under current law. That said, there was evidence that the increase in benefits due was moderated by a 23 percent increase in the share of Stage 2 beneficiaries earning above the BOND yearly amount.

In addition to the full sample results, Gubits et al. (2018) examined the impacts of the benefit offset on seven subgroups defined by baseline characteristics--duration of SSDI receipt, employment at enrollment, Medicaid Buy-In availability, age, primary impairment of major affective disorder, primary impairment of back disorder, and any post-secondary education. The subgroup analysis found no evidence of a clear pattern of behavioral effects for any subgroup (Gubits et al. 2018). While race and ethnicity were not part of the subgroup analysis, the research team leveraged other available data to perform 364 tests of differences in impacts. There was one observation related to the windfall argument that we highlight here. Individuals employed at baseline had a larger increase in SSDI benefits, which is likely due to the larger potential windfall effect compared to those unemployed at baseline.

3.2. Estimation Approach

In this section, we analyze impacts of the benefit offset on benefits due and offset use for a new subgroup in the Stage 2 sample. This analysis examines whether the impact of the benefit offset on SSDI benefits due and offset use varies across self-reported racial or ethnic groups. These analyses complement the prior subgroup tests and align with SSA's priority of assessing the effects of its interventions across race and ethnicity. We approach this research question by considering a series of one-to-one comparisons rooted in the impact estimation methodology and existing data detailed in Section 2.2.3 of the Final Evaluation Report (Gubits et al. 2018).[7] A key distinction in this analysis is that we estimate the models that compare two racial or ethnic groups at a time. Thus, we calculate three sets of differentials: Black versus White, Black versus Hispanic, and Hispanic versus White. We use t-tests to determine whether the differentials are statistically significant and F tests to compare across subgroups. We consider four Stage 2 comparisons: T21 versus C2; T22 versus C2; T21 versus T22 ; and T21 + T22 versus C2.

[7] A detailed description of the empirical specification, corresponding justification, and included covariates of the Stage 2 impact methodology can be found in Volume 2 of the Final Evaluation Report.

3.3. Estimated Impacts on SSDI Benefits Due for Subgroups Defined by Self-Reported Race and Ethnicity

Our approach produced four sets of results for each of the three self-reported racial/ethnic group comparisons. Only the Black-White comparison had a clear pattern of evidence that a differential exists. As such, we limit our discussion to those relevant estimates (shown in Exhibit 3-1). The full set of analyses associated with Exhibit 3-1 are provided in the Appendix (Exhibits B-1 through B-4).

In the exploratory analysis of differential impacts on SSDI benefits due between Black and White beneficiaries, we find consistent evidence of a positive difference in estimated impact. Exhibit 3-1 reports the Black versus White differential in annual SSDI benefits due between 2012 and 2019 by assignment group comparison. As an example of how to interpret the estimates, the top left value of $304 means that the impact estimate for Black volunteers (i.e., the average difference between Black T21 and Black C2 volunteers) is $304 larger than the impact estimate for White volunteers (i.e., the average difference between White T21 and White C2 volunteers).

Exhibit 3-1. Estimated Impacts on SSDI Benefits Due for Black versus White Stage 2 Volunteers by Assignment Group Comparisons

Outcome	T21 vs C2	T22 vs C2	T22 vs T21	T21+T22 vs C2
Benefits due in 2012	$304 ($262)	$425 ($304)	$121 ($242)	$351 ($220)
Benefits due in 2013	$482 ($285)	$615* ($320)	$133 ($187)	$534** ($167)
Benefits due in 2014	$477 ($303)	$667* ($337)	$190 ($161)	$552** ($208)
Benefits due in 2015	$543 ($324)	$531 ($363)	$-12 ($259)	$538* ($254)
Benefits due in 2016	$655* ($334)	$669 ($378)	$14 ($270)	$661** ($227)
Benefits due in 2017	$686* ($348)	$995** ($393)	$309 ($343)	$808*** ($229)
Benefits due in 2018	$715* ($361)	$932** ($411)	$216 ($286)	$800** ($267)
Benefits due in 2019	$521 ($376)	$493 ($429)	$-27 ($336)	$510 ($300)

Notes: Analysis of SSA administrative records (from the MEF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey. See Chapter 2 of Gubits et al. (2018) for variable definitions. Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Impact estimates are regression-adjusted for baseline characteristics and are calculated as the difference between the within Black impact (e.g., T21 vs C2) and the within White impact (e.g., T21 vs C2). Unweighted sample sizes: Black T21 = 1,253, Black T22 = 792, Black C2 = 1,248, White T21 = 2,612, White T22 = 1,694, and White C2 = 2,612. */**/*** Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Comparing the estimates across columns of Exhibit 3-1, we observe two key patterns. First, we find a pattern of positive Black-White differentials in benefits due for Stage 2 volunteers in any treatment group relative to the control group. This relationship holds for the T21 versus C2, T22 versus C2, and T21+T22 versus C2 analyses. Second, the Black-White differential fluctuates across negative and positive values in the comparison of T22 and T21, and all those annual estimates are statistically insignificant. This pattern indicates there is no evidence of a Black-White differential between the two Stage 2 treatment groups.

In our earlier discussion of Exhibit 2-2, we noted that, at the time of study enrollment, a small proportion of Black Stage 2 volunteers were working, compared to White Stage 2 volunteers. Gubits et al. (2018) found evidence that individuals who were employed at baseline had larger SSDI benefit increases, which was likely driven by the windfall effect. The patterns in Exhibit 3-1 indicate a separate mechanism other than the windfall effect associated with baseline employment drives the Black-White differential in SSDI benefits due.

3.4. Estimated Impacts on Offset Usage for Subgroups Defined by Self-Reported Race and Ethnicity

The final evaluation of BOND found limited use of the benefit offset. Gubits et al. (2018) document that through 2016 utilization rates were 3.7 percent among Stage 1 treatment subjects, 15.8 percent among Stage 2 T21 subjects, and 15.4 percent among Stage 2 T22 subjects. They went on to show that, among Stage 2 volunteers, the probability of offset use increased for younger subjects, those with short benefit duration, those working at baseline, those with higher levels of education, and those who did not report fair or poor health at baseline.

We extend those exploratory analyses by considering the differentials with race and ethnicity. As in section 3.3, we restrict the presentation of our findings to the Black-White differential. Exhibit 3-2 reports the Black versus White differential in annual offset use between 2012 and 2019 by assignment group comparison—the Appendix (Exhibits B-5 through B-8) includes the full set of estimates associated with these results. Similar to the analysis of benefits due, we find some evidence of a positive Black-White differential impact on any offset use among treatment subjects (T21 and T22 combined) relative to the control subjects. The pattern emerges for the demonstration years 2017 through 2019. There is no evidence of a Black-White differential impact of EWIC versus WIC (T22 versus T21).

Exhibit 3-2. Estimated Impacts on Any Offset Use for Black versus White Stage 2 Volunteers by Assignment Group Comparisons

Outcome	T21 vs C2	T22 vs C2	T22 vs T21	T21+T22 vs C2
Any offset use in 2012	0.4% (0.9%)	0.1% (1.1%)	0.3% (1.4%)	0.3% (0.6%)
Any offset use in 2013	-0.2% (1.3%)	-0.5% (1.4%)	0.3% (1.7%)	0.3% (1.0%)
Any offset use in 2014	0.0% (1.1%)	-0.8% (1.3%)	0.9% (1.0%)	0.3% (0.8%)
Any offset use in 2015	-0.4% (1.4%)	0.7% (2.3%)	1.0% (1.7%)	0.1% (1.6%)
Any offset use in 2016	- 0.1% (1.3%)	0.8% (1.5%)	0.9% (1.3%)	0.2% (1.2%)
Any offset use in 2017	2.0% (1.3%)	1.8% (1.3%)	0.2% (1.4%)	1.9%* (1.0%)
Any offset use in 2018	2.9%** (1.1%)	1.1% (1.0%)	1.8% (1.2%)	2.2%** (0.8%)
Any offset use in 2019	1.1% (0.7%)	1.6%* (0.8%)	0.5% (1.0%)	1.3%* (0.6%)

Notes: Analysis of SSA administrative records (from the MEF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey. See Chapter 2 of Gubits et al. (2018) for variable definitions. Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Impact estimates are regression-adjusted for baseline characteristics and are calculated

as the difference between the within Black impact (e.g., T21 vs C2) and the within White impact (e.g., T21 vs C2). Unweighted sample sizes: Black T21 = 1,253, Black T22 = 792, Black C2 = 1,248, White T21 = 2,612, White T22 = 1,694, and White C2 = 2,612.
*/**/*** Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

3.5. Discussion of Differential Effects of Benefit Offset

Sections 3.3 and 3.4 present findings of positive Black-White differentials in benefits due in years 2013 through 2018 and offset usage in years 2017 through 2019. The lack of corresponding estimated impacts in the same years for these two outcomes necessitated further examination. The principals of intersectional analysis call for researchers to consider the other factors, associated with race and ethnicity, that may contextualize evidence of differentials (Collins 1991; Crenshaw 1989).

As first step in this contextualization, we examine the percent of the sample with any benefits due. Exhibit 3-3 summarizes the findings from this subsequent set of analyses (see Exhibits B-9 though B-12 for full tables). We find that the positive Black-White differential for treatment subjects (T21 and T22 combined) relative to the control subjects holds for the outcome of percent of sample with any benefits due. We do not find evidence of a differential impact for EWIC versus WIC (T22 versus T21).

Exhibit 3-3. Estimated Impacts on Any Benefits Due Use for Black versus White Stage 2 Volunteers by Assignment Group Comparisons

Outcome	T21 vs C2	T22 vs C2	T22 vs T21	T21+T22 vs C2
Any benefits due in 2012	1.4% (1.3%)	2.1% (1.5%)	0.7% (0.9%)	1.7%* (0.8%)
Any benefits due in 2013	1.7% (1.6%)	2.5% (1.7%)	0.8% (1.0%)	2.0%** (0.9%)
Any benefits due in 2014	2.7% (1.7%)	3.3% (1.8%)	0.6% (1.3%)	2.9%** (1.0%)
Any benefits due in 2015	2.3% (1.8%)	2.8% (2.0%)	0.5% (1.9%)	2.5%** (0.8%)
Any benefits due in 2016	2.9% (2.0%)	3.7% (2.2%)	0.7% (1.8%)	3.2%** (1.3%)
Any benefits due in 2017	3.3% (2.2%)	4.4% (2.4%)	1.1% (2.2%)	3.7%** (1.2%)
Any benefits due in 2018	2.7% (2.3%)	3.4% (2.5%)	0.7% (2.2%)	3.0%* (1.3%)
Any benefits due in 2019	1.4% (1.3%)	2.1% (1.5%)	0.7% (0.9%)	1.7%* (0.8%)

Source: Analysis of SSA administrative records (from the DAF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.
Notes: Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics. Unweighted sample sizes: Black T21 = 1,253, Black T22 = 792, Black C2 = 1,248, White T21 = 2,612, White T22 = 1,694, and White C2 = 2,612.
*/**/*** Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

As we strive to understand what is driving the observed differentials, we find it helpful to consider how one can reorganize the differential's components. The Black-White differential in this empirical framework can be written as

$$(B_T - B_C) - (W_T - W_C) \qquad \text{(Eq. 1)}$$

where *B* identifies the estimated average outcomes for Black subjects in the treatment group, *T*, and control group, *C*. Likewise, *W* identifies the outcomes for White subjects. The differential as specified in Equation 1 compares the Black treatment impact against the White treatment impact. We will label these differences as "within race." Notably, we can rewrite Equation 1 into the following form.

$$(B_T - W_T) - (B_C - W_C) \quad \text{(Eq. 2)}$$

This new form shows that the same differential estimate can be viewed as comparing the Black-White treatment impact to the Black-White control impact. We will label these differences as "within assignment group." Thus, the positive Black-White differentials observed in sections 3.3 and 3.4 could be due to any combination of within race and within assignment group differences.

Examination of the detailed Appendix tables shows that the differences in benefits due appears to be driven by changes in Black and White control group subjects over time. For example, Exhibit B-9 and B-10 demonstrate that the average for Black control subjects is lower in every year than it is for White control subjects (i.e., more Black control subjects are losing their benefits every year than White control subjects). This finding aligns with the finding that Black volunteers are younger (on average) than White volunteers and were more likely to be working at baseline. Likewise, the differential in amount of benefits due (Exhibits B-4) reveals a reduction in benefits due by 17.0 percent for Black treatment subjects, 19.1 percent for Black control subjects, 16.3 percent for White treatment subjects, and 16.5 percent for White control subjects. Thus, again we find lower levels of benefits due for Black control subjects as well as a faster rate of decline.

In an attempt to understand these patterns for Black control subjects, we examine potential factors like termination of entitlement. We find no evidence of a Black-White control group differential on medical termination. We observe that termination-due-to-work is higher for Black control subjects than for White control subjects in 2013, 2014, 2015, and 2016 – with statistically significant differences in 2015 and 2016. This finding reinforces the contextualization that age and employment at baseline, which differed by race, influence our estimates of a differential impact. A benefit of our empirical specification, which ran separate regressions by race, is that it allows for the role of all regressors to vary by race. This specification also prevents the inclusion of a regressor (e.g., age) from accounting for the role of the regressor *across* races. Hence, it is possible for racial differences in age and employment at baseline to influence the estimated differential, despite those underlying measures being in our model.

4. Predicting Race and Ethnicity of Stage 2 Volunteers and Non-Volunteers

In Chapter 2, we discussed intersectional variations seen across the Stage 2 volunteers by comparing key baseline characteristics across race, ethnicity and sex. We found five themes of variations including demographic dimensions, education gaps, employment differentials, self-reported health, and limitations to work. Now we compare Stage 2 volunteers to the non-volunteers as well as to the nationally representative Stage 1 sample. But because we don't have complete self-reported race and ethnicity for Stage 1 or the Stage 2 non-volunteers, we test a procedure that leverages name and geocode information to predict race and ethnicity. After demonstrating the quality of the estimates from the imputation algorithm, we argue that SSA could use this approach to predict race and ethnicity and to address some of the needs identified in the Equity Action Plan. It is important to note that this chapter is not an assessment or evaluation of BOND. Rather, we use BOND data to showcase how a specific imputation approach can be used to estimate race and ethnicity among the SSDI population.

4.1. Defining the BIFSG Methodology

The Office of the Assistant Secretary for Planning and Evaluation and the Centers for Medicare and Medicaid Services have funded research supporting the development and advancement of statistical procedures of the BIFSG to predict race and ethnicity of enrollees in the health insurance marketplace (Sorbero, Euller, Kofner, and Elliott 2022). The Institute of Medicine of the National Academies recommends the use of indirect estimation, including the BIFSG-family of algorithms, of race and ethnicity when self-reported data are unavailable (Nerenz, McFadden, and Ulmer 2009). The Abt team obtained the algorithm from publicly available documentation, made available by the research team at RAND. The BIFSG approach is deeply rooted in the healthcare space, where it has been tested as a tool to predict race and ethnicity using healthcare data. The methodology started using only surname and geocode information (McCaffrey and Elliott 2008; Elliott, Fremont, Morrison, Pantoja, and Lurie 2008). Since its origin, it has advanced to include first name as part of the prediction data (Voicu 2018). The premise of the BIFSG approach is that one can use the first name and surname of an individual to predict their race and ethnicity based on the race and ethnicity of neighbors within their Census block group. Leveraging both name data and geocode data significantly improves the accuracy of the estimation method compared to methods that only use one data source (Elliott et al. 2009).[8] The BIFSG method has been widely tested and reviewed.[9]

---

[8] As described in Voicu (2018), the BIFSG approach estimates

$$p(r|s,f,g) = \frac{p(r|s) \cdot p(f|r) \cdot p(g|r)}{\sum_{r=1}^{6} p(r|s) \cdot p(f|r) \cdot p(g|r)}$$

where "$p(r|s,f,g)$ is the updated (posterior) probability of being of race\ethnicity *r*, given surname *s*, first name *f*, and geographic area *g*; $p(r|s)$ is the probability that a person is of race/ethnicity *r*, given that the person has surname *s*…; $p(f|r)$ is the probability that a person has first name *f*, given that the person is of race/ethnicity *r*…; $p(g|r)$ is the probability that a person resides in geographic area *g*, given that the person is of race/ethnicity *r*."

[9] A robust discussion of the BIFSG method and related publications can be found at https://www.rand.org/health-care/tools-methods/bisg.html.

The BIFSG succeeds the BISG, which is an earlier version of the algorithm that did not aim to predict race and ethnicity using first names. This enhancement leverages the Tzioumis data, which relate first name to race and ethnicity based on mortgage data used in Tzioumis (2018). Given that we were unsure of how well a sample of mortgage applicants would represent the SSDI population, we used the optional calibration feature of the BIFSG. In this context, calibration refers to using additional data linking first name to race and ethnicity from a subset of the study sample. We are able to calibrate the BIFSG estimates for the Stage 2 non-volunteers using the first names and self-reported race and ethnicity of the Stage 2 volunteers.[10] We use first name data from the Stage 2 volunteers because we think the relationship between their first names and race and ethnicity is more likely to be similar to Stage 2 non-volunteers than is the sample from mortgage applicants. As we will highlight in the results section, the calibration step improved the accuracy of our predictions.

While the BIFSG approach provides valuable insights into race and ethnicity, users of the estimates must understand how they differ from self-reported data. Typical self-reported race and ethnicity data allow individuals to provide nuanced information. For example, one might see Black Non-Hispanic individuals, Black Hispanic individuals, Asian and White individuals, and many other intersections of race and ethnicity within a single dataset. The BIFSG approach assigns each individual a probability of identifying into any one of the following groups: Hispanic, Black, Asian Pacific Islander (API), American Indian/Alaska Native (AIAN), White, and Multiracial. In total, these six probabilities sum to one. Moreover, a pioneer in the development of the BISG-family of algorithms made this important distinction between the predicted probabilities and self-reported data for the healthcare analysis space:

> Despite their improvement in accuracy, indirect methods such as BISG cannot replace the information gained from self-reported data. However, they are fast and not resource intensive. Given the multi-year time horizon for health plans (and others) to collect enough information to be actionable, and the strong desire of health plans and others to take action to address racial/ethnic disparities in care, such methods can serve as a bridge until the time when adequate self-reported data are available. (Elliott et al. 2009, pp. 81)

Another important distinction about the BIFSG methodology is how the approach differs from missing data imputation techniques. The BIFSG methodology aims to predict race and ethnicity when those variables are entirely missing from the data. While the accuracy of the BIFSG can be improved by calibrating the results with some self-reported race and ethnicity data, the calibration step is optional and the only required variables are name and address. Alternatively, there exists a robust set of imputation techniques (e.g., Multiple Imputation by Chained Equations, hot-deck, and regression-based single imputation) to address missing values among variables within a data set. The appropriate missing value imputation approach varies given the believed reason for the missing values and the amount of missing cells (Jakobsen, Gluud, Wetterslev, and Winkel 2017). Given the differences in the approach and purpose of the BIFSG methodology, the missing data thresholds that guide missing value imputation techniques do not apply to the BIFSG algorithm. The observed missing value rates and variation in the level of information recorded, even for the cases with available race and ethnicity data in current SSA records,

[10] The calibration step fits a multinomial logistic regression of self-reported race and ethnicity on the six uncalibrated race and ethnicity probabilities. It then produces predictions (i.e., calibrated probabilities) for each of the race and ethnicity groups using the regression results.

complicate the use of missing-value imputation approaches but do not restrict the use of the BIFSG methodology.

4.2. Data Used in the Assessment of the BIFSG Methodology

To assess the BIFSG methodology for the SSA use case, we incorporate data from multiple sources, which are documented in Exhibit 4-1. We begin by describing the creation of the BIFSG probabilities. The Abt team identified the BIFSG as a well-reviewed race and ethnicity prediction algorithm and contacted the RAND Corporation (RAND) to request the command files. RAND shared the BISFG algorithm, which is in the public domain and freely available for use subject to terms and conditions. Because the prediction methodology is available to the public, a wide range of researchers are using the algorithm in different circumstances, building evidence about its results and accuracy.

As part of the package provided by RAND, we received the BIFSG algorithm script files and all supporting documents. These documents include county names and FIPS codes, race and ethnicity shares by 2010 Census block group, race and ethnicity shares by 2010 Census tract, race and ethnicity 2010 Census counts at the national level, surname match files, and the "Tzioumis first name" file. The input data into the BIFSG algorithm come from the BOND sample data used to conduct Stage 1 random assignment, taken from SSA's Master Beneficiary Record (MBR) prior to random assignment. The variables are first name, surname, and geocoded Census block group of home address. These three variables are available for the Stage 2 volunteers and non-volunteers. We use the self-reported race and ethnicity data from the Stage 2 survey of T21, T22, and C2 to calibrate the base BIFSG probabilities.

Exhibit 4-1. Sources of Data in Assessment of BIFSG Methodology

Measure	Source	Notes
BIFSG estimation		
Name (beneficiary)	MBR	
Address (beneficiary)	MBR	
Self-report race and ethnicity (beneficiary)	Stage 2 Baseline Survey; Stage 1 36-Month Survey	
BIFSG supporting datasets	RAND	2010 Census extracts, Tzioumis data on first names
Individual characteristics		
Age at recruitment	MBR	Non-volunteer age is calculated at the midpoint of the Stage 2 recruitment period, December 2011
Sex	MBR	
Years of DI receipt	MBR	
Concurrent status	MBR	

Note: MBR refers to SSA's Master Beneficiary Record. To form the BOND sample, in late 2010 SSA extracted administrative records from the MBR for all SSDI beneficiaries in the 10 BOND sites who were between ages 20 and 59 as of May 2011, were receiving benefits based on disability, and who were not part of another SSA demonstration. See Gubits et al. (2018) for more information on the BOND Stage 1 and Stage 2 samples. RAND refers to the RAND Corporation.

We examine differential race and ethnicity patterns for volunteers and non-volunteers using by individual characteristics. We use measures taken from the Master Beneficiary Record: DOB, sex, years of SSDI receipt (as of May 2011), and concurrent status (as of May 2011). While additional individual characteristics are available for volunteers from the Stage 2 baseline survey, we prioritize comparisons to non-volunteers.

4.3. Viability of BIFSG Methodology

The possibility for SSA to use its existing data to predict the race and ethnicity of its applicants and beneficiaries presents an opportunity to assess and monitor equitable access of its programs and services.[11] In this section of the report, we demonstrate the viability of the BIFSG methodology by assessing how it performs relative to self-reported data. Then, we examine differentials in race and ethnicity by BOND assignment group and consider how individual characteristics intersect with these patterns.

In the analyses presented within this section, we make seven key observations:

- The BIFSG methodology is a viable approach for SSA to use to predict race and ethnicity among the SSDI population for whom SSA does not currently hold race and ethnicity data and is improved by a within sample calibration to correct for bias driven by the differences between the SSDI population and the overall U.S. population;
- The calibrated BIFSG performs better than the BIFSG and the BIFSG performs better than the BISG in predicting race and ethnicity in the BOND Stage 2 volunteers.
- The accuracy of the calibrated BIFSG is sufficiently high for the four testable groups—Black (excellent), Hispanic (excellent), White (strong/excellent), and Asian Pacific Islander (acceptable);
- Stage 2 volunteers are more likely to be Black and less likely to be White than the non-volunteers in the solicitation pool;
- The higher proportions of Black women to Black men as well as the lower proportions of Hispanic and White women to Hispanic and White men among Stage 2 volunteers seen in the intersectional analysis holds for the Stage 2 non-volunteers;
- There is evidence of variation in short duration status and age among racial and ethnic groups, even for the Stage 2 non-volunteers; and
- Even after accounting for a set of individual characteristics, there exists a statistical relationship between multiple BIFSG estimates and measures of sex and the decision to volunteer.

4.3.1. Examination of the Accuracy of the BIFSG Methodology Against Self-Reported Data

The literature examining the potential use and accuracy of the BISG-family of algorithms focuses on three methods to describe the accuracy of the estimates. Earlier work compared the relative efficiency of the estimates, looking at the relationships of the squared correlation between predicted and self-reported race and ethnicity (Elliott et al. 2008; Elliott et al. 2009). A less common approach uses false negative rates and false positive rates between the self-reported race and ethnicity and the maximum a posteriori classification of the predictions, which assigns an individual to a single race/ethnicity based on their

[11] SSA's EAP states that to, "to reduce barriers in [its] administrative system, [it] will increase collection of race and ethnicity data to determine whether [its] programs are equitably serving [its] applicants and beneficiaries, revise [its] policies and practices to expand options for service delivery, ensure equitable access to unrepresented claimants in the disability application process, decrease burdens for people who identify as gender diverse or transgender in the Social Security Number card application process, and increase access to [its] research grant programs for Historically Black Colleges and Universities and Minority Serving Institutions and procurement opportunities for small and disadvantaged businesses" (page 2).

largest probability (Voicu, 2018). Recent work has focused on concordance statistics of the area under the curve analysis derived from receiver operating characteristic curves (Adjaye-Gbewonyo et al. 2014; Sartin et al. 2021). In this report, we use the area under the curve approach to assess accuracy.

To assess the accuracy of the BIFSG methodology for SSDI beneficiaries, we first examine the relative shares of each racial and ethnic group when considering the self-reported data and three different estimates under the BIFSG approach. The top panel of Exhibit 4-2 shows the summary of self-reported groupings among those that self-reported (i.e., the Stage 2 volunteers). The three most represented identities are White (57.47 percent), Black (25.44 percent), and Hispanic (9.30 percent). The next largest group were those that identified as Multiracial (5.86 percent), but this group represents just over half of the third largest group. The next three rows report the BISG uncalibrated estimates (note this method does not use the first name), BIFSG uncalibrated estimates, and the BIFSG calibrated estimates.

Each progressive version of the algorithm shifts the estimates closer to the self-reported data. Without the first name, the uncalibrated BISG underestimates the share of Black participants and overestimates the share the White participants. Including the first name in the uncalibrated BIFSG estimates slightly increases the share of Blacks but incorrectly decreases the share of Hispanics and increases the share of Whites. The BIFSG calibrated estimates leverage the self-reported identities in our data to adjust the initial BISG and BIFSG estimates. Specifically, we see a large increase in the shares of Blacks and Hispanics and a decrease in the share of Whites. Evidence presented in Exhibit 4-3 indicates that the calibration improves the accuracy of the predictions for those with self-reported race and ethnicity data in both Stages 1 and 2.

We find similar patterns in the racial and ethnic composition when the predictions include first names and calibration techniques for individuals who do not have self-reported data. Prior to the calibration in the final BIFSG approach, the predicted identities had a considerably smaller share of Black participants and a higher share of White participants. The addition of the calibration phase redistributed the White share into the Multiracial and Black identities.

Exhibit 4-2. Imputation Analysis Comparing Imputed Vectors by Self-Reported Status

Estimate Type	Black	Hispanic	White	API	AIAN	Multiracial
Self-reported (Volunteers)						
Self-Report	25.44	9.30	57.47	0.95	0.97	5.86
BISG Uncalibrated	22.64	9.89	63.09	1.17	0.76	2.46
BIFSG Uncalibrated	23.39	8.76	63.72	1.02	0.71	2.39
BIFSG Calibrated	25.52	9.32	57.40	0.96	0.97	5.84
Did not self-report (Non-Volunteers)						
Self-Report	.	.	.	.	.	.
BISG Uncalibrated	18.43	9.66	66.85	1.33	0.80	2.94
BIFSG Uncalibrated	18.91	8.54	67.68	1.20	0.78	2.89
BIFSG Calibrated	21.20	9.81	61.17	1.10	1.03	5.88
All (Volunteers + Non-Volunteers)						
BISG Uncalibrated	18.51	9.66	66.78	1.32	0.80	2.93
BIFSG Uncalibrated	19.00	8.54	67.61	1.20	0.77	2.88
BIFSG Calibrated	21.28	9.60	61.10	1.10	1.03	5.88

Source: SSA administrative records (MBR), Stage 2 baseline survey, and BIFSG supplemental data.
Note: Imputation analysis of Stage 2 volunteers and non-volunteers. Cells are the average probability, expressed between 0 and 100, of the racial-ethnic group by type. Thus, all rows sum to 100—allotting for rounding error.

Next, we examine the accuracy of the BIFSG by considering both the Stage 1 participants and the Stage 2 volunteers with self-reported data. The calibrated BIFSG approach successfully classifies BOND Stage 2 volunteers that identify as Black, Hispanic, or White. Exhibit 4-2 demonstrates that the approach creates average percentages that closely match the self-reported data relative to the uncalibrated estimates. Exhibit 4-3 further shows that both the uncalibrated and calibrated BIFSG are able to effectively differentiate between racial and ethnic groups. We conducted an area under the curve analyses by stage grouping (i.e., Stage 2 volunteers with self-reported race and ethnicity data only, Stage 1 participants with self-reported race and ethnicity data, and the combination of those groups) for each of the identities except American Indian/Alaska Native and Multiracial due to data limitations. Concordance statistics for each logistic regression of the binary indicator of the identity group on the continuous measure for that group are reported in Exhibit 4-3. Hosmer and Lemeshow (2000) recommend the following minimum values as thresholds for interpreting concordance statistics: 0.7 is acceptable, 0.8 is strong, and 0.9 is excellent. By this standard, the calibrated BIFSG approach is excellent at discerning Black and Hispanic individuals in all three groupings and is on the borderline of strong and excellent at discerning White individuals. The concordance statistic for the Asian Pacific Islander group is firmly in the acceptable range for the Stage 2 only group and the combined group but is in the strong range from the Stage 1 group. Given the substantially lower proportion of Stage 2 volunteers and Stage 1 participants who self-identify as Asian Pacific Islander, we are not surprised by the lower classification score for this sub-group. Exhibit 4-3 also demonstrates that the accuracy of the methods increase with the incorporation of additional information in the model for nearly all race/ethnicities and groups, with the exception of the BIFSG uncalibrated versus BIFSG calibrated Black estimates for all three groups and the BIFSG uncalibrated versus BIFSG calibrated API estimate for the Stage 1 group, which all had relatively small reductions in the concordance statistic. These patterns indicate that the BIFSG calibrated approach produces the most accurate predictions.

Finally, the comparable levels of accuracy seen among the Stage 2 volunteers, who had a high rate of available self-reported race and ethnicity data, with the less than one percent of Stage 1 participants with self-reported race and ethnicity data from the 36-month survey suggest that the calibrated BIFSG approach can work well for the SSDI population and the accuracy seen in this report is not solely a function of features of the Stage 2 volunteers tied to their selection to participate in BOND. The calibration leveraged self-reported data from Stage 1 participants and Stage 2 volunteers to produce a single prediction framework for both samples. The comparable levels of accuracy seen in Exhibit 4-3 indicate that the evidence that the BIFSG is a viable tool for SSA is not dependent on unique features of the Stage 2 volunteers. For the Stage 2 volunteers (i.e., those with self-reported data), we expected that the calibration process would converge the BIFSG and self-reported estimates because we are calibrating to information we know to hold true for that group. A key finding from this analysis is that it confirms the power of using self-reported race and ethnicity data on a small proportion of the sample to improve the prediction of race and ethnicity. Hence, the calibration served as a bias correction relative to the uncalibrated BIFSG approach, where the bias results from the underlying differences between the SSDI population and the overall U.S. population.

Exhibit 4-3. Assessment of the Accuracy of the BIFSG Calibration

Calibration Status	Black	Hispanic	White	API
Stage 2 Volunteers				
BISG Uncalibrated	0.9373958	0.9049046	0.8795144	0.7237336
BIFSG Uncalibrated	0.9422295	0.9183915	0.8929283	0.7386709
BIFSG Calibrated	0.9396362	0.9355987	0.8947989	0.7750035
Stage 1 Self-Reported Race and Ethnicity				
BISG Uncalibrated	0.9460650	0.8813721	0.8969588	0.8005518
BIFSG Uncalibrated	0.9485833	0.9013451	0.9093248	0.8389370
BIFSG Calibrated	0.9472224	0.9161974	0.9103049	0.8310379
All				
BISG Uncalibrated	0.9394236	0.8979524	0.8838953	0.7481315
BIFSG Uncalibrated	0.9436871	0.9132939	0.8968396	0.7707082
BIFSG Calibrated	0.9414189	0.9301407	0.8984457	0.7929234

Source: SSA administrative records (MBR), Stage 2 baseline survey, Stage 1 36-month survey, and BIFSG supplemental data.
Note: Concordance statistics are reported by identity group.

4.3.2. Examination of Differentials in Race and Ethnicity Probabilities by BOND Assignment Groups

In this section of the report, we study the differentials in BIFSG probabilities by BOND assignment groups and individual characteristics. We explore this topic to inform our understanding of the underlying patterns in racial and ethnic identities by these individual characteristics. One common thread observed in these analyses is that the Stage 2 non-volunteers appear more similar to the Stage 1 sample than to the Stage 2 volunteers in terms Black and White representation, as measured by the calibrated BIFSG.

The differential shares of Black and White beneficiaries across Stage 1 and Stage 2 participants are demonstrated in Exhibit 4-4. After applying sample weights to make the averages nationally representative, we see a roughly six percentage point gap in the Black probability across Stage 1 (22-24 percent) and Stage 2 (29 percent).[12] Likewise, the share of White beneficiaries in Stage 1 (59-60 percent) is approximately 5 percentage points higher than the share in Stage 2 (54-55 percent). We find similar proportions of other racial identity groups in Stage 1 and Stage 2.

Exhibit 4-4. BIFSG Probabilities by Assignment Group Using BOND Administrative Weights

Assignment Group	Black	Hispanic	White	API	AIAN	Multiracial
Stage 1						
C1	22.31	9.19	60.39	1.15	1.04	5.92
T1	23.72	9.69	58.53	1.21	1.03	5.82
Stage 1 Combined	23.13	9.48	59.30	1.19	3.08	2.35
Stage 2						
C2	28.72	9.09	54.36	0.99	0.96	5.88
T21	28.81	8.87	54.29	1.13	0.97	5.91
T22	29.12	8.37	54.68	1.02	0.88	5.94

[12] We used the relevant Stage 2 administrative weights and Stage 2 administrative weights as used in the BOND final evaluation (Gubits et al. 2018). For a full description of the weights, see section B.1.6 and B.2.6 of Volume 2 of the Final Report (Gubits et al. 2018).

Assignment Group	Black	Hispanic	White	API	AIAN	Multiracial
Stage 2 Combined	28.85	8.84	54.41	1.05	0.94	5.91
p-value	*0.000*	*0.001*	*0.000*	*0.005*	*0.000*	*0.011*

Source: SSA administrative records (MBR), Stage 2 baseline survey, and BIFSG supplemental data.
Note: Reported probabilities come from the calibrated BIFSG algorithm. This analysis includes Stage 1 and Stage 2 beneficiaries. We do not include Stage 2 non-volunteers because they do not have corresponding administrative weights given that they did not enroll in BOND. To test the equality of the Stage 1 combined sample against the Stage 2 combined sample, we implemented t tests and report the corresponding p-values.

Focusing only on the probabilities among Stage 2 assignment groups reveals that Stage 2 volunteers were more likely to be Black and less likely to be White than the non-volunteers in the solicitation pool. Similar to the weighted estimates in Exhibit 4-4, we see a roughly five to six percentage point volunteer versus non-volunteer gap for Black and White probabilities in Exhibit 4-5. The tests of equality of the BIFSG estimates within racial and ethnic groups show that there are statistically significant differences at the conventional levels of five percent for the Black, White, and Asian Pacific Islander probabilities. Along with the approximately three percentage point shift for White and Black share between the weighted and unweighted probabilities for C2, T21, and T22, the evidence suggests that the racial representation of non-volunteers is more comparable to that of the Stage 1 participants.

Exhibit 4-5. BIFSG Probabilities by Stage 2 Assignment Groups

Assignment Group	Black	Hispanic	White	API	AIAN	Multiracial
C2	25.81	9.81	56.62	0.95	1.00	5.81
T21	25.78	9.30	57.08	0.97	1.02	5.85
T22	26.55	8.88	56.82	0.97	0.90	5.89
Non-volunteers	20.55	9.62	61.84	1.12	1.03	5.83
p-value	0.000	0.223	0.000	0.030	0.095	0.463

Source: SSA administrative records (MBR), Stage 2 baseline survey, and BIFSG supplemental data.
Note: F tests assess the equality of estimates within racial and ethnic groups, and the corresponding p-values are reported.

The BIFSG probabilities also reveal sex differences for Black, Hispanic, and White Stage 2 beneficiaries. Exhibit 4-6 shows that for most assignment groups in Stage 2 there are higher proportions of Black women than Black men but the reverse patterns for Hispanic and White beneficiaries. This relationship holds for volunteers, excluding Hispanic beneficiaries in T21 and White beneficiaries in T22, and non-volunteers but is the largest for Black beneficiaries in both levels and relative terms. The higher proportions of Black women to Black men as well as the lower proportions of Hispanic and White women to Hispanic and White men among Stage 2 volunteers seen in the intersectional analysis holds for the Stage 2 non-volunteers.

Exhibit 4-6. BIFSG Probabilities by Stage 2 Assignment Groups and Sex

	Sex	C2	T21	T22	Non-Volunteers
Black	**Male**	23.69	22.63	24.83	19.68
	Female	27.91	28.81	28.23	21.54
	p-value	0.000	0.000	0.006	0.000
Hispanic	**Male**	10.90	9.71	10.08	10.24
	Female	8.73	8.91	7.69	8.92
	p-value	0.002	0.248	0.005	0.000

	Sex	C2	T21	T22	Non-Volunteers
White	Male	57.82	59.84	57.44	62.16
	Female	55.42	54.44	56.21	61.48
	p-value	0.021	0.000	0.348	0.000
API	Male	0.99	0.97	0.95	1.13
	Female	0.92	0.97	0.98	1.11
	p-value	0.617	0.981	0.866	0.370
AIAN	Male	0.91	1.10	0.92	1.02
	Female	1.10	0.94	0.88	1.04
	p-value	0.041	0.070	0.504	0.246
Multiracial	Male	5.69	5.76	5.78	5.76
	Female	5.92	5.94	6.00	5.92
	p-value	0.001	0.008	0.011	0.000

Source: SSA administrative records (MBR), Stage 2 baseline survey, and BIFSG supplemental data.
Note: Average BIFSG probability reported by sex and assignment group. We conduct t tests of the equality of probabilities by sex within assignment group and report corresponding p-values.

We find evidence of a relationship between the length of time receiving SSDI benefits (at BOND baseline) and race/ethnicity. In Exhibit 4-7, we present the BIFSG probabilities by assignment group and by short duration status. Short duration is defined as less than or equal to 36 months of benefit receipt at BOND baseline, approximately May 2011. The BOND sampling purposely oversampled short duration beneficiaries so across the total sample it has a higher proportion of short duration to longer duration beneficiaries. To demonstrate how to interpret the findings in Exhibit 4-7, the top left estimate means 26.40 percent of C2 beneficiaries that did not have short duration status are Black.

By examining the data by race and ethnicity, we see variation in racial and ethnic composition by short duration status beneficiaries across the BIFSG groups. There is no evidence of differences in the proportion of Black or American Indian/Alaska Native beneficiaries across the two short duration statuses. Among beneficiaries with short duration status, we find higher proportions of Hispanic (in all assignment groups) and Asian Pacific Islander (in C2 and non-volunteer groups) beneficiaries. Conversely, we observe lower proportions of White (in C2, T22, and non-volunteer groups) and Multiracial (in all groups) beneficiaries among those with short duration status.

Exhibit 4-7. BIFSG Probabilities by Stage 2 Assignment Groups and Short Duration Status

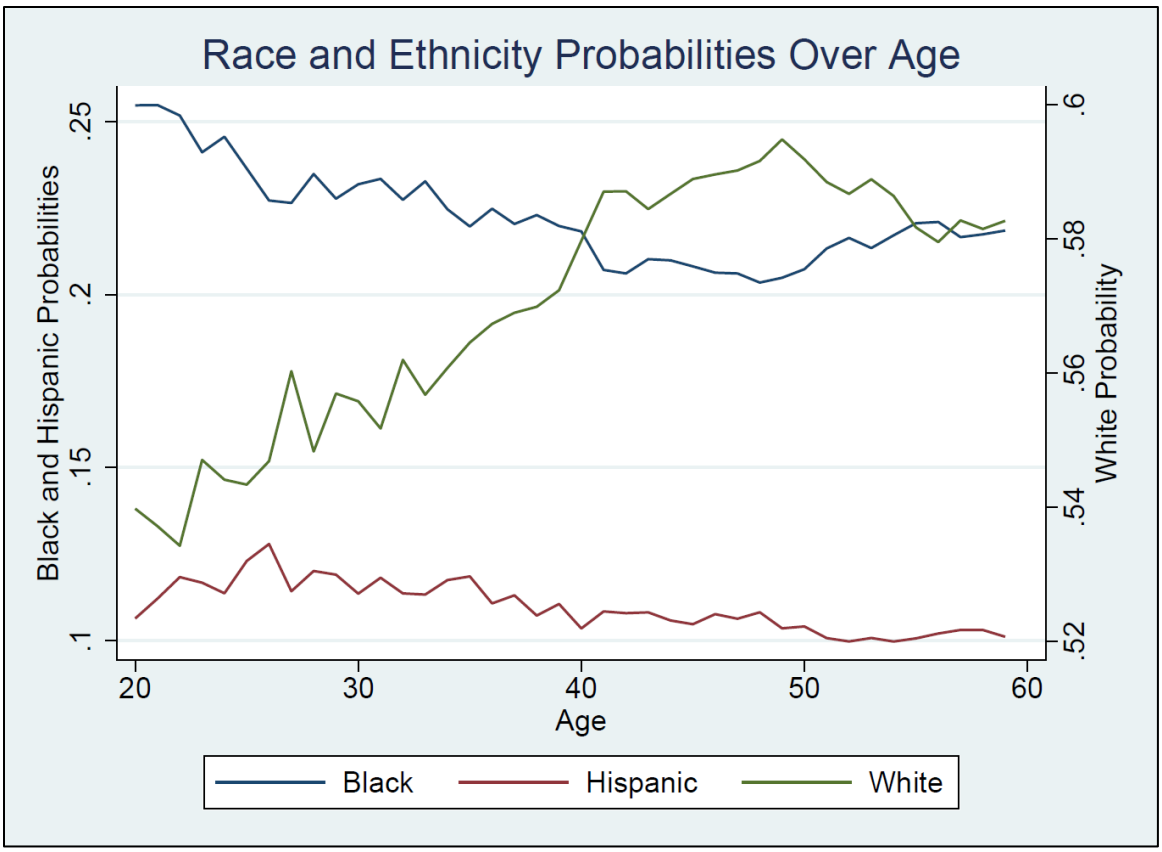
	Short Duration	C2	T21	T22	Non-volunteers
Black	No	26.40	26.00	25.83	20.64
	Yes	25.48	25.66	26.97	20.48
	p-value	0.360	0.733	0.371	0.206
Hispanic	No	7.89	8.41	7.28	8.70
	Yes	10.89	9.80	9.82	10.45
	p-value	0.000	0.053	0.004	0.000
White	No	58.07	57.60	59.18	62.71
	Yes	55.80	56.80	55.42	61.06
	p-value	0.036	0.454	0.005	0.000
API	No	0.79	1.07	0.85	1.01
	Yes	1.04	0.92	1.04	1.22
	p-value	0.071	0.304	0.286	0.000

	Short Duration	C2	T21	T22	Non-volunteers
AIAN	No	0.91	0.98	0.87	1.03
	Yes	1.06	1.04	0.92	1.03
	p-value	0.129	0.502	0.419	0.963
Multiracial	No	5.94	5.95	5.99	5.91
	Yes	5.73	5.79	5.83	5.77
	p-value	0.004	0.029	0.077	0.000

Source: SSA administrative records (MBR), Stage 2 baseline survey, and BIFSG supplemental data.
Note: Average BIFSG probability reported by short duration status and assignment group. We conduct t tests of the equality of probabilities by short duration status within assignment group and report corresponding p-values.

Next we examine the relationship between age and BIFSG estimates. We document differing BIFSG probabilities by age for Black, Hispanic, and White groups. Exhibit 4-8 shows a strong positive relationship between age and the probability of being White, which means that the BIFSG algorithm predicts that the Stage 2 sample is relatively more White for older beneficiaries relative to younger ones. Conversely, the race-age relationship for Black and Hispanic probabilities is negative.

Exhibit 4-8. Black, Hispanic, and White BIFSG Probabilities Over Age



Source: SSA administrative records (MBR), Stage 2 baseline survey, and BIFSG supplemental data.
Note: The Black (blue line) and Hispanic (red line) probabilities are plotted against the left vertical axis. The White probability (green line) is plotted against the right vertical axis.

4.3.3. Relationship between BIFSG Estimates, Individual Characteristics, and Assignment Group

In this section of the report, we aim to examine if and how assignment group and individual factors influence the BIFSG predicted race. We use a regression framework to assess the relative roles that each of the measures plays on the BIFSG race estimate. This analysis aims to sharpen our understanding of the statistical relationship of these factors on the BIFSG estimates within a multivariate framework that accounts for the correlated factors. Specifically, we estimate a series of unweighted ordinary linear regressions of the form:

$$P_i = \alpha + X_i\beta + Z_i\gamma + \epsilon_i$$

where P_i is the probability of a specific race for individual i. X_i is a set of individual characteristics including a quadratic in age, sex, and a quadratic in years of SSDI receipt. Z_i is a set of indicators equaling one that identify person i's BOND assignment group, with C2 as the omitted group. Finally, ϵ_i captures the remaining idiosyncratic error.

We are examining the relationship between the BIFSG estimates and individual characteristics within the inclusion of assignment group. The estimates in Exhibit 4-9 statistically confirm three differences within the Stage 2 sample seen in the earlier BIFSG analysis. First, the negative convex relationship between age and the probability of being Black, and a positive concave relationship with age and the probability of being White. Second, we find a positive and statistically significant estimate of female on the probability of being Black and a negative and statistically significant estimate of it on the probability of being Hispanic or White. Finally, while the relationship of years of SSDI receipt is negative and convex for the probability Black, Hispanic, and Asian Pacific Islanders, it is positive and concave for the probability of White.

Exhibit 4-9. Relationship between BIFSG Estimates and Individual Characteristics among Stage 2 Sample

Covariate	Black	Hispanic	White	API	AIAN	Multiracial
Age	-0.00441***	0.000255	0.00424***	-0.000216	0.0000109	0.000118*
	(-7.29)	(0.52)	(6.13)	(-1.83)	(0.18)	(2.50)
Age Squared	0.0000443***	-0.00000555	-0.0000395***	0.00000201	-0.000000162	-0.00000105*
	(6.73)	(-1.04)	(-5.25)	(1.56)	(-0.25)	(-2.05)
Female	0.0208***	-0.0140***	-0.00828***	-0.000288	0.000134	0.00160***
	(17.04)	(-14.16)	(-5.93)	(-1.21)	(1.09)	(16.79)
Years of DI	-0.000654**	-0.00220***	0.00293***	-0.000241***	0.00000536	0.000162***
	(-2.90)	(-12.11)	(11.37)	(-5.47)	(0.24)	(9.18)
Years of DI Squared	0.0000289***	0.0000293***	-0.0000588***	0.00000303	-0.000000231	-0.00000213***
	(3.52)	(4.43)	(-6.28)	(1.90)	(-0.28)	(-3.33)
Constant	0.306***	0.115***	0.497***	0.0181***	0.0101***	0.0536***
	(22.81)	(10.62)	(32.38)	(6.89)	(7.49)	(51.17)
Observations	*238,848*	*238,848*	*238,848*	*238,848*	*238,848*	*238,848*
R-squared	*0.002*	*0.003*	*0.002*	*0.000*	*0.000*	*0.002*

Source: SSA administrative records (MBR), Stage 2 baseline survey, and BIFSG supplemental data.
Note: Regression estimates and standard errors in parentheses are reported.
*/**/*** Estimate is significantly different from zero at the .05/.01/.001 levels, respectively, using a two-tailed t-test.

As demonstrated in Exhibit 4-10, we extend the analysis in Exhibit 4-9 to include assignment group. The estimates for individual characteristics remain fairly stable after including assignment group, which aligns with earlier estimates demonstrating the equality of baseline characteristics across assignment groups (Gubits et al. 2013). Notably, we find large statistically significant estimates for non-volunteers relative to C2 but insignificant estimates for either of the treatment groups. Exhibit 4-10 shows that after accounting for a set of individual characteristics, there remains a positive statistical relationship between volunteering for Stage 2 and being Black. We also find a negative relationship between volunteering for Stage 2 and being White or Asian Pacific Islander.

Exhibit 4-10. Relationship between BIFSG Estimates, Assignment Group, and Individual Characteristics among Stage 2 Sample

Covariate	Black	Hispanic	White	API	AIAN	Multiracial
Assignment Group						
T21	-0.000504	-0.00497	0.00479	0.000175	0.000122	0.000391
	(-0.08)	(-1.02)	(0.69)	(0.15)	(0.20)	(0.83)
T22	0.00732	-0.00887	0.00158	0.000197	-0.00104	0.000810
	(1.06)	(-1.59)	(0.20)	(0.15)	(-1.50)	(1.50)
Non-volunteers	-0.0518***	0.00130	0.0481***	0.00207*	0.000275	0.0000628
	(-11.94)	(0.37)	(9.69)	(2.44)	(0.63)	(0.19)
Individual Characteristic						
Age	-0.00471***	0.000285	0.00450***	-0.000205	0.0000137	0.000116*
	(-7.79)	(0.58)	(6.51)	(-1.73)	(0.23)	(2.47)
Age Squared	0.0000486***	-0.00000597	-0.0000433***	0.00000185	-0.000000201	-0.00000103*
	(7.39)	(-1.12)	(-5.74)	(1.43)	(-0.31)	(-2.00)
Female	0.0204***	-0.0139***	-0.00798***	-0.000276	0.000137	0.00160***
	(16.76)	(-14.13)	(-5.72)	(-1.16)	(1.12)	(16.76)
Years of DI	-0.000470*	-0.00222***	0.00277***	-0.000248***	0.00000375	0.000162***
	(-2.08)	(-12.19)	(10.75)	(-5.62)	(0.17)	(9.22)
Years of DI Squared	0.0000249**	0.0000297***	-0.0000554***	0.00000318*	-0.000000195	-0.00000215***
	(3.04)	(4.49)	(-5.92)	(1.99)	(-0.24)	(-3.36)
Constant	0.358***	0.114***	0.449***	0.0160***	0.00979***	0.0536***
	(25.41)	(9.98)	(27.82)	(5.80)	(6.93)	(48.62)
Observations	*238,848*	*238,848*	*238,848*	*238,848*	*238,848*	*238,848*
R-squared	*0.003*	*0.003*	*0.003*	*0.001*	*0.000*	*0.002*

Source: SSA administrative records (MBR), Stage 2 baseline survey, and BIFSG supplemental data.
Note: Regression estimates and standard errors in parentheses are reported.
*/**/*** Estimate is significantly different from zero at the .05/.01/.001 levels, respectively, using a two-tailed t-test.

Finally, we consider the role of sex on the relationship of assignment group and individual characteristics on BIFSG estimates. Given the evidence within this report of sex intersecting with multiple other individual factors, we ran each of the regressions separately by sex. Thus, the six BIFSG categories and two sex categories required twelve regressions, with corresponding estimates reported in Exhibits 4-11 and 4-12. The key finding from this set of analyses highlights an important composition variation between the non-volunteers and volunteers. After accounting for the role of individual factors on the BIFSG estimate, volunteering for Stage 2 continues to have a positive relationship with the BIFSG

estimates for Black women and men—with a more positive role for Black women. Similarly, volunteering for Stage 2 has a negative relationship with the BIFSG probabilities for White men and women.

Exhibit 4-11. Relationship between Black, Hispanic, and White BIFSG Estimates and Assignment Group and Individual Characteristics among Stage 2 Sample by Sex

Assignment Group	Black		Hispanic		White	
Sex	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Assignment Group						
T21	-0.0109	0.00902	-0.0118	0.00187	0.0203*	-0.00987
	(-1.29)	(1.04)	(-1.63)	(0.29)	(2.06)	(-1.02)
T22	0.0111	0.00360	-0.00789	-0.00975	-0.00388	0.00692
	(1.16)	(0.36)	(-0.95)	(-1.31)	(-0.35)	(0.62)
Non-volunteers	-0.0411***	-0.0627***	-0.00255	0.00509	0.0402***	0.0563***
	(-6.85)	(-10.00)	(-0.49)	(1.09)	(5.72)	(8.04)
Individual Characteristic						
Age	-0.00312***	-0.00676***	0.00184**	-0.00163*	0.00137	0.00843***
	(-3.91)	(-7.33)	(2.67)	(-2.37)	(1.47)	(8.19)
Age Squared	0.0000297***	0.0000730***	-0.0000228**	0.0000146	-0.00000761	-0.0000881***
	(3.40)	(7.29)	(-3.03)	(1.95)	(-0.75)	(-7.88)
Years of DI	0.000772*	-0.00184***	-0.00201***	-0.00249***	0.00128***	0.00447***
	(2.56)	(-5.43)	(-7.73)	(-9.83)	(3.61)	(11.82)
Years of DI Squared	0.00000879	0.0000341**	0.0000185*	0.0000451***	-0.0000271*	-0.0000810***
	(0.82)	(2.71)	(1.99)	(4.79)	(-2.15)	(-5.77)
Constant	0.311***	0.436***	0.0829***	0.139***	0.527***	0.344***
	(16.74)	(20.24)	(5.16)	(8.65)	(24.17)	(14.34)
Observations	*126,370*	*112,478*	*126,370*	*112,478*	*126,370*	*112,478*
R-squared	*0.002*	*0.004*	*0.002*	*0.002*	*0.001*	*0.005*

Source: SSA administrative records (MBR), Stage 2 baseline survey, and BIFSG supplemental data.
Note: Regression estimates and standard errors in parentheses are reported.
*/**/*** Estimate is significantly different from zero at the .05/.01/.001 levels, respectively, using a two-tailed t-test.

Exhibit 4-12. Relationship between API, AIAN, and Multiracial BIFSG Estimates and Assignment Group and Individual Characteristics among Stage 2 Sample by Sex

Assignment Group	API		AIAN		Multiracial	
Sex	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Assignment Group						
T21	-0.000152	0.000488	0.00192*	-0.00163	0.000647	0.000127
	(-0.09)	(0.30)	(2.26)	(-1.88)	(0.95)	(0.19)
T22	-0.000313	0.000689	0.000125	-0.00218*	0.000882	0.000732
	(-0.16)	(0.37)	(0.13)	(-2.21)	(1.14)	(0.97)
Non-volunteers	0.00190	0.00221	0.00114	-0.000579	0.000417	-0.000294
	(1.55)	(1.90)	(1.88)	(-0.93)	(0.86)	(-0.62)

Assignment Group	API		AIAN		Multiracial	
Sex	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Individual Characteristic						
Age	-0.000228	-0.000167	0.000106	-0.000105	0.0000213	0.000233***
	(-1.40)	(-0.97)	(1.32)	(-1.15)	(0.33)	(3.34)
Age Squared	0.00000189	0.00000168	-0.00000113	0.000000990	-2.86e-08	-0.00000225**
	(1.06)	(0.90)	(-1.28)	(0.99)	(-0.04)	(-2.97)
Years of DI	-0.000225***	-0.000276***	0.0000159	-0.00000945	0.000174***	0.000150***
	(-3.65)	(-4.39)	(0.52)	(-0.28)	(7.18)	(5.85)
Years of DI Squared	0.00000271	0.00000374	-0.000000574	0.000000200	-0.00000225**	-0.00000215*
	(1.23)	(1.60)	(-0.53)	(0.16)	(-2.59)	(-2.26)
Constant	0.0171***	0.0142***	0.00664***	0.0137***	0.0553***	0.0529***
	(4.49)	(3.55)	(3.53)	(6.40)	(36.92)	(32.47)
Observations	*126,370*	*112,478*	*126,370*	*112,478*	*126,370*	*112,478*
R-squared	*0.001*	*0.001*	*0.000*	*0.000*	*0.001*	*0.001*

Source: SSA administrative records (MBR), Stage 2 baseline survey, and BIFSG supplemental data.
Note: Regression estimates and standard errors in parentheses are reported.
*/**/*** Estimate is significantly different from zero at the .05/.01/.001 levels, respectively, using a two-tailed t-test.

5. Conclusion

As SSA examines equitable access to its programs and services for all beneficiaries, the agency must have information on the racial and ethnic diversity in its program participants and on the barriers faced by beneficiaries of different racial and ethnic groups. If SSA is to center equity in its demonstrations and research, as called for in its Equity Action Plan, it is essential that SSA take full advantage of available data from its demonstrations to examine diversity in program participants. This analysis offers a more detailed description of the SSDI beneficiaries who volunteered for Stage 2 of BOND than previously available, examining how experiences before volunteering for BOND vary by race, ethnicity, and sex. While we do not seek to, and are not able to, conclude why those variations exist at baseline (i.e., racism, structural, systemic inequities), an essential first step towards an assessment of equity is to use the data available as fully as possible.

SSA might use the results of this analysis to inform the design of SSA services and future interventions in a way that centers on equity. For example, the results might indicate demographic dimensions that SSA could use for stratification and subgroup analyses. While the BOND Stage 2 sample does not represent all SSDI beneficiaries, it is aligned with a group of beneficiaries eager to explore employment-focused options and work incentives to improve their economic wellbeing, which is a highly relevant group as SSA considers future demonstrations and interventions.

Four key differences in the BOND Stage 2 volunteers stand out as likely immediate gaps for SSA to consider. First, we find clear differences in the shares of those employed at the time they volunteered for BOND, from 16.4 to 32.8 percent, for Black men and White women, respectively. Second, race/ethnicity, sex, and their intersection all point to different average educational patterns for beneficiaries. We find that White beneficiaries have higher average educational attainment than their racial and ethnic peers and that women, on average, have more schooling than men. The strongest differentials exist within the Black and White communities between men and women. SSA's service plans and future interventions must be cognizant that the racial and ethnic employment and education gaps observed in the general population are present within this sample of beneficiaries, which may call for tailored services based on the needs of historically marginalized groups. Third, at the time they volunteered for BOND, each racial identity group varied in terms of the likelihood of reporting perceived limitations as barriers to employment. Interventions that aim to reduce barriers will need to be responsive to the different needs (e.g., reliable transportation, childcare, etc.) across these identity groups. Finally, workplace accommodations are utilized differentially among the groups. To the extent that accommodations are a feature of future interventions, SSA will need to engage in proactive outreach to the subgroups with the most limited experience with accommodations.

Analysis of the differential impact of the BOND benefit offset intervention finds evidence of a positive Black-White impact differential. We observe this differential impact on the amount of benefits due, percent of beneficiaries with benefits due, and offset use. After considering the *within race* and *within assignment group* differences that lead into the estimated differential, we identify that levels and changes for Black control group subjects relative to White control groups subjects drive the overall findings. We attribute this pattern to differences, at baseline, in age and employment that themselves vary by race.

The evidence from the assessment of the BIFSG methodology demonstrates that this name and geocode-based algorithm is a viable race and ethnicity prediction tool for SSA. Existing administrative data

contain the necessary information to implement the methodology. Our analysis of the BIFSG approach indicates that calibrating estimates using internal self-reported race and ethnicity data improves the accuracy of the predictions relative to the standard BIFSG methodology. Supplemental self-reported race and ethnicity data can be pulled from demonstration surveys and other sources available to SSA, so it can appropriately calibrate the BIFSG estimates. Moreover, we find evidence of similarities and differences in characteristics between the Stage 2 volunteers and non-volunteers: volunteers are more likely to be Black and less likely to be White than non-volunteers; and both volunteers and non-volunteers have higher proportions of Black women than Black men as well as lower proportions of Hispanic and White women than Hispanic and White men. We also demonstrate that a differential take-up rate from the solicitation pool into the volunteer sample holds even after accounting for a set of individual characteristics. These observed differential volunteering rates provide SSA with a more complete picture of the characteristics of SSDI beneficiaries willing to participate in a benefit offset program and the systemic barriers that interventions must recognize to equitably serve all beneficiaries.

6. Implications and Next Steps

Even if SSA provides services to all beneficiaries equally, this does not mean that beneficiaries have equal access. This disparity exists because each person faces unique barriers to accessing SSA services and work opportunities. Because of systemic racism, sexism, and other forms of discrimination, these barriers likely differ systemically along a person's race, ethnicity, sex, and other demographics. Environmental factors such as access to education, transportation, healthcare, and childcare further exacerbate the systemic differences in a beneficiary's experiences navigating the social security system.

In a recent summary of the lessons learned across SSA demonstrations, Nichols and Hemmeter (2021) highlighted that among the reviewed demonstrations none analyzed if race or ethnicity influenced enrollment or service delivery. They went on to argue that improving our understanding of the racial and ethnic differences in participation and outcomes of SSDI and SSI beneficiaries is aligned with the President's Executive Order on Advancing Racial Equity and Support for Underserved Communities through the Federal Government and a pressing need for future research. They underscore the need to examine root causes of variations observed as well as determine whether and how recruitment and outreach might lead to disparities.

This report directly addresses SSA's priority to identify inequalities in their programs and services while also providing a more detailed snapshot of the BOND Stage 2 volunteers. Moreover, it lays the foundation for future equity-focused analysis using SSA data by demonstrating that the calibrated BIFSG is a viable algorithm to predict race and ethnicity. While we show that the BIFSG estimates are significantly improved after being calibrated with some self-reported race and ethnicity data, this requirement is within scope for SSA given its going efforts to encourage self-reporting by applicants.

Should SSA use race and ethnicity prediction algorithms, like the BIFSG, to assess equity within its programs and demonstrations, it will need to consider the ethical implications of these analyses (Kaplan and Bennet 2003; Brown et al. 2021). Sartin et al. (2021) argue that race and ethnicity are social constructs; therefore, analyses that examine outcomes by race and ethnicity must be cautious about inadvertently attributing disparities as being caused by race and ethnicity. Instead, they encourage researchers to consider potential roles created by the social and environmental constructs associated with race and ethnicity. More specifically to equity-focused analyses using imputed data, Brown et al. (2021) outline a set of recommendations for policy-focused research on when and how to impute race and ethnicity as well as ethical and empathy-related risks, including inaccurate representation, reidentification after opting out, and failing to fully recognize the personhood of individuals in the analysis.

By leveraging existing data and implementing ethical standards on imputation, SSA can achieve a significant aspect of its Equity Action Plan. This data science innovation will enable SSA to conduct equity analyses for several relevant populations, including claimants, beneficiaries, and demonstration participants; thus, allowing SSA to better understand who is seeking benefits, who is getting benefits, and how can demonstrations inform future services. In addition to this new information to SSA, the agency will also be able to produce stratified reporting to educate the public about these race and ethnicity

differentials. The Centers for Medicare and Medicaid Services has developed a comparable knowledge dissemination approach.[13]

[13] Information on the stratified reporting at the Centers for Medicare and Medicaid Services can be found at https://www.cms.gov/About-CMS/Agency-Information/OMH/research-and-data/statistics-and-data/stratified-reporting.

References

Adjaye-Gbewonyo, Robert A. Bednarczyk, Robert L. Davis, and Saad B. Omer. (2014). Using the Bayesian Improved Surname Geocoding Method (BISG) to create a working classification of race and ethnicity in a diverse managed care population: a validation study. *Health Services Research*, *49(1)*, 268-283.

Bauer, Greta R. and Ayden I. Scheim. (2019). Methods for analytic intercategorical intersectionality in quantitative research: discrimination as a mediator of health inequalities. *Social Science & Medicine*, *226*, 236-245.

Brown, K. Steven, LesLeigh Ford, Shena Ashley, Alena Stern, and Ajjit Narayanan 2021. *Ethics and Empathy in Using Imputation to Disaggregate Data for Racial Equity: Recommendations and Standards Guide.* Washington, D.C.: Urban Institute.

Collins, Patricia H., 1991. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Routledge, New York.

Crenshaw, Kimberle. 1989. *Demarginalizing the Intersection of Race and Sex: a Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, vol. 1. University of Chicago Legal Forum. Issue, pp. 139–167 Article 8.

Elliott, Marc N., Allen Fremont, Peter A. Morrison, Philip Pantoja, and Nicole Lurie. 2008. A new method for estimating race/ethnicity and associated disparities where administrative records lack self-reported race/ethnicity. *Health services research*, *43*(5p1), pp.1722-1736.

Gubits, Daniel, Rachel Cook, Stephen Bell, Michelle Derr, Jillian Berk, Ann Person, David Stapleton, Denise Hoffman, and David Wittenburg. 2013. *BOND Implementation and Evaluation: Stage 2 Early Assessment Report.* Rockville, MD: Abt Associates.

Gubits, Daniel, Judy Geyer, Denise Hoffman, Sarah Croake, Utsav Kattel, David Judkins, Stephen Bell, and David Stapleton. 2017. *BOND Implementation and Evaluation: 2015 Stage 2 Interim Process, Participation, and Impact Report.* Report for Social Security Administration, Office of Program Development & Research. Cambridge, MA: Abt Associates; and Washington, DC: Mathematica Policy Research.

Gubits, Daniel R., Judy Geyer, David Stapleton, David Greenberg, Stephen Bell, Austin Nichols, Michelle Wood, Andrew McGuirk, Denise Hoffman, Meg Carroll, Sarah Croake, Utsav Kattel, David R Mann, and David Judkins. 2018a. *BOND Implementation and Evaluation: Final Evaluation Report,* Vol. 1. Report for the Social Security Administration. Cambridge, MA: Abt Associates; and Washington, DC: Mathematica Policy Research.

Gubits, Daniel R., Judy Geyer, David Stapleton, David Greenberg, Stephen Bell, Austin Nichols, Michelle Wood, Andrew McGuirk, Denise Hoffman, Meg Carroll, Sarah Croake, Utsav Kattel, David Mann, and David Judkins. 2018b. *BOND Implementation and Evaluation: Final Evaluation Report.* Vol. 2, *Technical Appendices*. Report for Social Security Administration. Cambridge, MA: Abt Associates; and Washington, DC: Mathematica Policy Research.

Hosmer, David W. and Stanley Lemeshow. 2000. "Applied Logistic Regression, Second Edition." John Wiley & Sons, Inc.

Janus C. Jakobsen, Christian Gluud, Jorn Wetterslev, and Per Winkel. 2003. When and How Should Multiple Imputation be Used for Handling Missing Data in Randomized Clinical Trials – A Practical Guide with Flowcharts. *BMC Medical Research Methodology*, *17*(1), pp.1-10.

Kaplan, Judith B. and Trude Bennett. 2003. Use of race and ethnicity in biomedical publication. *Jama*, *289*(20), pp.2709-2716.

Martin, Patricia P., Why Researchers Now Rely on Surveys for Race Data on OASDI and SSI Programs: A Comparison of Four Major Surveys (January 22, 2016). Research and Statistics Note, No. 2016-01, January 2016, Available at SSRN: https://ssrn.com/abstract=2725921

Martin, Patricia P., and John L. Murphy, 2014. "African Americans: Description of Social Security and Supplemental Security Income Participation and Benefit Levels Using the American Community Survey." Research and Statistics Note No. 2014-01. https://www.socialsecurity.gov/policy/docs/rsnotes/rsn2014-01.html.

McCaffrey, Daniel F. and Marc N. Elliott. 2008. Power of tests for a dichotomous independent variable measured with error. *Health services research*, *43*(3), pp.1085-1101.

Nerenz, David R., Bernadette McFadden, and Cheryl Ulmer, eds. 2009. "Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement." Institute of Medicine. The National Academies Press, Washington D.C.

Nichols, Austin and Jeffrey Hemmeter. 2021. "An Introduction to Disability Policy and SSA's Demonstrations." In Lessons from SSA Demonstrations for Disability Policy and Future Research, edited by Austin Nichols, Jeffrey Hemmeter, and Debra Goetz Engler, 1–32. Rockville, MD: Abt Press

Sartin, Emma B., Kristina B. Metzger, Melissa R. Pfeiffer, Rachel K. Myers, and Allison E. Curry. 2021. Facilitating research on racial and ethnic disparities and inequities in transportation: Application and evaluation of the Bayesian Improved Surname Geocoding (BISG) algorithm. *Traffic injury prevention*, *22 (supplement 1)*, pp.S32-S37.

Sorbero, Melony E., Roald Euller, Aaron Kofner, and Marc N. Elliott. 2022. *Imputation of Race and Ethnicity in Health Insurance Marketplace Enrollment Data, 2015–2022 Open Enrollment Periods*. https://www.rand.org/content/dam/rand/pubs/research_reports/RRA1800/RRA1853-1/RAND_RRA1853-1.pdf.

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 (contract deliverable 2.2 under Contract SS00-10-60011), Office of Program Development & Research. Cambridge, MA: Abt Associates, December 2010.

Social Security Administration. 2022. "Social Security's Equity Action Plan in Agreement with EO 13985." https://www.ssa.gov/open/materials/SSA-EO-13985-Equity-Action-Plan.pdf.

Tzioumis, Konstantinos. 2018. "Demographic Aspects of First Names." *Scientific Data* 5.1: 1-9.

Voicu, Ioan (2018) Using First Name Information to Improve Race and Ethnicity Classification, Statistics and Public Policy, 5:1, 1-13, DOI: 10.1080/2330443X.2018.1427012

Whitebread, Geoffrey, Stephanie Dolamore, and Brendan Stern. 2022. Quantitative Intersectionality: Imperatives and Opportunities for Advancing Social Equity. *Public Administration Review 1– 13.* https://doi.org/10.1111/puar.13555

Appendix A: Intersectional Analysis

Exhibit A-1. Baseline Demographic Characteristics by Self-Reported Race and Ethnicity

Characteristic	All	Black	Hispanic	White	p-value	
Number of Beneficiaries	11360	3293	1140	6927		
Age						
Age	47.34	47.08	46.9	47.55	0.196724	
Age category						
20-29	5.57	5.22	5.75	5.73	0.163395	
30-39	14.61	16.17	15.81	13.54		
40-44	12.31	12.82	13.66	11.81		
45-49	17.35	16.95	16.82	17.67		
50-54	23.87	23.55	23.34	24.14		
55-60	26.28	25.29	24.61	27.11		
Marital Status						
Married	30.26	21.46	35.34	34.32	0	***
Widowed, Divorced, or Separated	38.15	41.2	39.02	36.31		
Never married	31.59	37.34	25.64	29.37		
Currently living with spouse or partner	36.35	26.73	42.12	40.75	0	***
Language Spoken at Home						
Primary language at home is not English	4.39	1.63	32.44	1.22	0.200041	
Educational Attainment						
High school degree or more	89.02	85.37	77.68	92.96	0.000432	***
College degree or more (incl. Associate, professional, Bachelor's)	32.69	26.22	24.55	37.68	0	***
Bachelor's degree or more	17.26	12.34	11.8	20.92	0.000006	***
Currently In School						
Currently enrolled in school or taking classes	7.9	9.64	7.95	6.93	0.012436	**
Currently working toward degree, certificate, or license	6.69	8.24	6.38	5.87	0.018312	**
Full-time student	3.43	4.57	3.37	2.79	0.01101	**
Part-time student	4.31	4.93	4.5	3.94	0.0672	*

Source: SSA administrative records (MBR) and Stage 2 baseline survey.

Note: This analysis compares mean differences race/ethnicity. We test the hypothesis that the mean characteristic is equal across race and ethnicity groups. If the characteristic is measured by a continuous or binary variable, we use a t-test to compare means and obtain p-values. If the characteristic is measured by a categorical variable, we use a Chi-Square test to compare the distributions and obtain p-values. Stars indicate statistical significance. One star indicates that the characteristics are statistically significantly different at a 10 percent significance level. Two stars indicate a significance level of 5 percent, and three stars indicates a significance level of 1 percent.

Exhibit A-2. Baseline Employment-Related Characteristics by Self-Reported Race and Ethnicity

Characteristic	All	Black	Hispanic	White	p-value	
Number of Beneficiaries	11360	3293	1140	6927		
Labor Force Participation						
Currently working at a job	26.16	18.84	20.69	31.16	0.000699	***
Currently looking for work	30.25	29.47	27.44	31.32	0.288122	
Of those not working, # of months since last worked (Median)	40.07	39.56	38.88	40.69	0.257282	
Job characteristics (for those currently working)						
Hours worked per week (Median)	19.68	22.68	21.2	19.49	0.005056	***
Tenure (in months) at current job (Median)	15.04	12.17	12.62	16.71	0.000455	***
Annual Earnings						
$0	74.04	80.94	79.01	69.25	0	***
$1-2,999	5.7	3.99	4.45	6.9		
$3,000-5,999	4.33	2.99	3.92	5.17		
$6,000-8,999	4.35	2.69	4.07	5.34		
$9,000-11,999	4.76	3.36	3.61	5.74		
$12,000-14,999	2.42	1.85	1.21	2.94		
$15,000 or above	4.41	4.17	3.73	4.66		
Work Accommodations						
Use of special equipment related to disability at work	6.38	4.58	4.26	7.74	0.0029	***
Use of personal assistance service at work	2.12	1.07	1.6	2.8	0.009495	***
Ability to Work						
Has physical or mental condition that limits ability to work	83.03	83.21	61.46	85.92	0.704113	
Had someone help with baseline interview	0.43	0.2	0.44	0.56	0.078937	*
Knowledge of Program Rules						
Ever heard of trial work period (TWP)	74.37	71.8	57.05	78.72	0.046821	**
Ever heard of extended period of eligibility (EPE)	21.86	19.21	14.9	24.53	0.00658	***

Source: SSA administrative records (MBR) and Stage 2 baseline survey.

Note: This analysis compares mean differences race/ethnicity. We test the hypothesis that the mean characteristic is equal across race and ethnicity groups. If the characteristic is measured by a continuous or binary variable, we use a t-test to compare means and obtain p-values. If the characteristic is measured by a categorical variable, we use a Chi-Square test to compare the distributions and obtain p-values. Stars indicate statistical significance. One star indicates that the characteristics are statistically significantly different at a 10 percent significance level. Two stars indicate a significance level of 5 percent, and three stars indicates a significance level of 1 percent.

Exhibit A-3. Baseline Health- and Transportation-Related Characteristics by Self-Reported Race and Ethnicity

Characteristic	All	Black	Hispanic	White	p-value
Number of Beneficiaries	11360	3293	1140	6927	
Self-reported Health Status					
Good, very good, or excellent	37.53	33.22	35.03	40.36	0.002225 ***
Hospital Use					
Stayed overnight in hospital in last 12 months	30.24	32.53	32.48	28.58	0.015154 **
Number of nights in hospital in last 12 months (Median)	4.15	4.01	3.96	4.28	0.109457
Health Insurance					
Have health insurance	94.54	93.31	93.39	95.42	0.005989 ***
Limitations (Agree or strongly agree responses)					
Inability to work because of a physical or mental condition	85.15	81.39	84.54	87.35	0.000723 ***
Inability to work because I do not have reliable transportation to and from work	18.15	22.3	20.6	15.42	0.000125 ***
Inability to work because I am caring for children or others	9.6	11.53	10.45	8.36	0.008751 ***
Difficult to work because I am afraid I will lose my disability benefits	40.71	37.24	46.97	41.61	0.211363
Inability to work because I am finishing a school or training program	4.32	4.75	4.39	4.07	0.422417
Workplaces are not accessible to people with my disability	43.86	46.05	49.6	41.67	0.049379 **
Lacking skills or training I need to return to work	33.17	33.06	37.69	32.47	0.63777
Difficulty to re-qualify for Social Security disability benefits in the future if I work	39.63	39.07	43.26	39.32	0.999467
Personal goals include moving up in a job or learning new job skills	88.88	91.77	90.33	87	0.000856 ***
Usual Mode of Transportation					
Own car, truck, or van	68.78	56.78	63.95	76.29	0.000248 ***
Public transportation	27.2	41.49	32.44	18.34	0.000747 ***
Friends or relatives	52.61	60.87	58.78	46.96	0.000039 ***
Walk	32.02	36.84	34.8	28.86	0.037353 **
Taxi, van, or paratransit service	20.29	29.4	22.74	14.79	0.004075 ***
Wheel or motorized scooter	6.4	5.14	6.65	7.06	0.041711 **
Other	5.11	3.02	3.34	6.58	0.008629 ***
Ability to Drive					
Able to drive a car	82.66	79.1	79.55	85.17	0.046468 **
Have a valid driver's license	81	73.31	76.55	86.04	0.000113 ***
Access to a car that runs	88.25	81.28	86.86	91.78	0.000007 ***

Source: SSA administrative records (MBR) and Stage 2 baseline survey.

Note: This analysis compares mean differences race/ethnicity. We test the hypothesis that the mean characteristic is equal across race and ethnicity groups. If the characteristic is measured by a continuous or binary variable, we use a t-test to compare means and obtain p-values. If the characteristic is measured by a categorical variable, we use a Chi-Square test to compare the distributions and obtain p-values. Stars indicate statistical significance. One star indicates that the characteristics are statistically significantly different at a 10 percent significance level. Two stars indicate a significance level of 5 percent, and three stars indicates a significance level of 1 percent.

Exhibit A-4. Baseline Demographic Characteristics by Sex

Characteristic	Women	Men	p-value	
Number of Beneficiaries	6434	6310		
Age				
Age	47.35	47.26	0.702217	
Age category				
20-29	4.84	6.35	0.008546	***
30-39	15.25	14.09		
40-44	12.91	11.9		
45-49	16.88	17.81		
50-54	24.4	23.19		
55-60	25.72	26.65		
Marital Status				
Married	27.55	32.49	0	***
Widowed, Divorced, or Separated	43.64	32.68		
Never married	28.81	34.83		
Currently living with spouse or partner	32.95	39.82	0.001005	***
Language Spoken at Home				
Primary language at home is not English	3.49	5.54	0.05024	*
Educational Attainment				
High school degree or more	90.99	87.16	0.003126	***
College degree or more (incl. Associate, professional, Bachelor's)	37.88	27.23	0	***
Bachelor's degree or more	20.25	14.18	0.000015	***
Currently In School				
Currently enrolled in school or taking classes	9.48	6.58	0.005642	***
Currently working toward degree, certificate, or license	8.02	5.41	0.001604	***
Full-time student	4.32	2.82	0.012932	**
Part-time student	4.93	3.67	0.042946	**

Source: SSA administrative records (MBR) and Stage 2 baseline survey.

Note: This analysis compares mean differences by sex. We test the hypothesis that the mean characteristic for women equals the mean characteristic for men. If the characteristic is measured by a continuous or binary variable, we use a t-test to compare means and obtain p-values. If the characteristic is measured by a categorical variable, we use a Chi-Square test to compare the distributions and obtain p-values. Stars indicate statistical significance. One star indicates that the characteristics are statistically significantly different at a 10 percent significance level. Two stars indicate a significance level of 5 percent, and three stars indicates a significance level of 1 percent.

Exhibit A-5. Baseline Employment-Related Characteristics by Sex

Characteristic	Women	Men	p-value	
Number of Beneficiaries	6434	6310		
Labor Force Participation				
Currently working at a job	26.96	23.97	0.002683	***
Currently looking for work	29.46	31.21	0.343149	
Of those not working, # of months since last worked (Median)	41.3	40.09	0.604539	
Job characteristics (for those currently working)				
Hours worked per week (Median)	19.47	19.91	0.003924	***
Tenure (in months) at current job (Median)	13.18	15.93	0.000036	***
Annual Earnings				
$0	73.33	76.03	0.001409	***
$1-2,999	6.6	4.84		
$3,000-5,999	4.35	4.02		
$6,000-8,999	4.72	3.75		
$9,000-11,999	4.35	4.72		
$12,000-14,999	2.4	2.22		
$15,000 or above	4.26	4.42		
Work Accommodations				
Use of special equipment related to disability at work	6.9	5.6	0.083699	*
Use of personal assistance service at work	2.11	2.16	0.868527	
Ability to Work				
Has physical or mental condition that limits ability to work	89.27	83.97	0.709433	
Had someone help with baseline interview	0.16	0.73	0.002176	***
Knowledge of Program Rules				
Ever heard of trial work period (TWP)	75.32	73.47	0.154	
Ever heard of extended period of eligibility (EPE)	22.78	20.97	0.008511	***

Source: SSA administrative records (MBR) and Stage 2 baseline survey.

Note: This analysis compares mean differences by sex. We test the hypothesis that the mean characteristic for women equals the mean characteristic for men. If the characteristic is measured by a continuous or binary variable, we use a t-test to compare means and obtain p-values. If the characteristic is measured by a categorical variable, we use a Chi-Square test to compare the distributions and obtain p-values. Stars indicate statistical significance. One star indicates that the characteristics are statistically significantly different at a 10 percent significance level. Two stars indicate a significance level of 5 percent, and three stars indicates a significance level of 1 percent.

Exhibit A-6. Baseline Health- and Transportation-Related Characteristics by Sex

Characteristic	Women	Men	p-value	
Number of Beneficiaries	6434	6310		
Self-reported Health Status				
Good, very good, or excellent	34.5	39.49	0.026295	**
Hospital Use				
Stayed overnight in hospital in last 12 months	31.05	30.04	0.248437	
Number of nights in hospital in last 12 months (Median)a	4.02	4.4	0.008921	***
Health Insurance				
Have health insurance	95.49	93.73	0.0002	***
Limitations (Agree or strongly agree responses)				
Inability to work because of a physical or mental condition	85.8	84.44	0.1708	
Inability to work because I do not have reliable transportation to and from work	17.43	20.47	0.000066	***
Inability to work because I am caring for children or others	12.22	7.29	0.000045	***
Difficult to work because I am afraid I will lose my disability benefits	40.4	41.37	0.501712	
Inability to work because I am finishing a school or training program	4.79	4.09	0.031959	**
Workplaces are not accessible to people with my disability	45.38	43.91	0.326903	
Lacking skills or training I need to return to work	32.86	34.68	0.128478	
Difficulty to re-qualify for Social Security disability benefits in the future if I work	40.13	39.45	0.38076	
Personal goals include moving up in a job or learning new job skills	88.23	89.6	0.05166	*
Usual Mode of Transportation				
Own car, truck, or van	70.19	66.46	0.004152	***
Public transportation	23.78	31.03	0.000017	***
Friends or relatives	56.77	49.62	0.000107	***
Walk	28.32	36.36	0.000001	***
Taxi, van, or paratransit service	21.66	19.44	0.100441	
Wheel or motorized scooter	5.96	6.91	0.134336	
Other	2.6	7.82	0.000038	***
Ability to Drive				
Able to drive a car	83.04	81.74	0.062779	*
Have a valid driver's license	83.59	77.94	0.000104	***
Access to a car that runs	88.24	87.57	0.356489	

Source: SSA administrative records (MBR) and Stage 2 baseline survey.

Note: This analysis compares mean differences by sex. We test the hypothesis that the mean characteristic for women equals the mean characteristic for men. If the characteristic is measured by a continuous or binary variable, we use a t-test to compare means and obtain p-values. If the characteristic is measured by a categorical variable, we use a Chi-Square test to compare the distributions and obtain p-values. Stars indicate statistical significance. One star indicates that the characteristics are statistically significantly different at a 10 percent significance level. Two stars indicate a significance level of 5 percent, and three stars indicates a significance level of 1 percent.

Appendix B: Estimated Impacts on SSDI Benefits Due and Offset Usage

Exhibit B-1. Estimated Impacts of the Offset Compared to Current Law (T21 Vs. C2) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and WIC (T21) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and WIC (T21) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Benefits due in 2012	$12,291	$11,837	$455* ($202)	$13,797	$13,646	$151 ($214)	$304 ($262)
Benefits due in 2013	$12,173	$11,368	$804*** ($222)	$13,532	$13,209	$323 ($212)	$482 ($285)
Benefits due in 2014	$11,944	$11,113	$832*** ($237)	$13,247	$12,892	$355* ($188)	$477 ($303)
Benefits due in 2015	$12,022	$10,928	$1,094*** ($255)	$13,306	$12,755	$551** ($204)	$543 ($324)
Benefits due in 2016	$11,689	$10,510	$1,179*** ($263)	$12,858	$12,334	$523** ($206)	$655† ($334)
Benefits due in 2017	$10,900	$9,891	$1,009*** ($279)	$12,156	$11,834	$323 ($244)	$686† ($348)
Benefits due in 2018	$10,392	$9,619	$773** ($286)	$11,602	$11,545	$58 ($308)	$715† ($361)
Benefits due in 2019	$10,208	$9,578	$630* ($298)	$11,500	$11,391	$109 ($340)	$521 ($376)

Source: Analysis of SSA administrative records (from the MEF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T21 = 1,253, Black C2 = 1,248, White T21 = 2,621, White C2 = 2,612.

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

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-2. Estimated Impacts of the Offset Plus WIC or EWIC Compared to Current Law (T22 Vs. C2) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and EWIC (T22) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and EWIC (T22) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Benefits due in 2012	$12,414	$11,837	$577** ($241)	$13,798	$13,646	$153 ($186)	$425 ($304)
Benefits due in 2013	$12,366	$11,368	$998*** ($253)	$13,591	$13,209	$382* ($196)	$615† ($320)
Benefits due in 2014	$12,183	$11,113	$1,070*** ($264)	$13,295	$12,892	$403* ($210)	$667† ($337)
Benefits due in 2015	$12,050	$10,928	$1,122*** ($287)	$13,347	$12,755	$591** ($224)	$531 ($363)
Benefits due in 2016	$11,676	$10,510	$1,166*** ($300)	$12,831	$12,334	$497* ($232)	$669 ($378)
Benefits due in 2017	$11,012	$9,891	$1,121*** ($308)	$11,959	$11,834	$126 ($243)	$995†† ($393)
Benefits due in 2018	$10,624	$9,619	$1,005** ($327)	$11,619	$11,545	$74 ($266)	$932†† ($411)
Benefits due in 2019	$10,298	$9,578	$719* ($342)	$11,617	$11,391	$226 ($315)	$493 ($429)

Source: Analysis of SSA administrative records (from the MEF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: See Chapter 2 of the Final Evaluation Report for variable definitions. Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T22 = 792 , Black C2 = 1,248, White T22 = 1,694 , White C2 = 2,612.

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

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-3. Estimated Impacts of the Offset Compared to Offset Plus WIC or EWIC (T21 Vs. T22) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and EWIC (T22) (1)	Average Outcome with Offset and WIC (T21) (2)	Impact Estimate (3)	Average Outcome with Offset and EWIC (T22) (4)	Average Outcome with Offset and WIC (T21) (5)	Impact Estimate (6)	
Benefits due in 2012	$12,414	$12,291	$122 ($214)	$13,798	$13,797	$2 ($209)	$121 ($242)
Benefits due in 2013	$12,366	$12,173	$193 ($209)	$13,591	$13,532	$60 ($191)	$133 ($187)
Benefits due in 2014	$12,183	$11,944	$239 ($272)	$13,295	$13,247	$48 ($188)	$190 ($161)
Benefits due in 2015	$12,050	$12,022	$28 ($300)	$13,347	$13,306	$40 ($216)	$-12 ($259)
Benefits due in 2016	$11,676	$11,689	$-13 ($294)	$12,831	$12,858	$-27 ($189)	$14 ($270)
Benefits due in 2017	$11,012	$10,900	$112 ($339)	$11,959	$12,156	$-197 ($220)	$309 ($343)
Benefits due in 2018	$10,624	$10,392	$232 ($234)	$11,619	$11,602	$16 ($236)	$216 ($286)
Benefits due in 2019	$10,298	$10,208	$89 ($308)	$11,617	$11,500	$117 ($264)	$-27 ($336)

Source: Analysis of SSA administrative records (from the MEF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: See Chapter 2 of the Final Evaluation Report for variable definitions. Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T21 = 1,253, Black T22 = 792, White T21 = 2,621, White T22 = 1,694.

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

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-4. Estimated Impacts of the Combined Offset Compared to Current Law (T21+T22 Vs. C2) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and WIC and EWIC (T22 + T21) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and WIC and EWIC (T22 + T21) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Benefits due in 2012	$12,340	$11,837	$503** ($184)	$13,797	$13,646	$151 ($148)	$351 ($220)
Benefits due in 2013	$12,248	$11,368	$880*** ($201)	$13,555	$13,209	$346* ($157)	$534†† ($167)
Benefits due in 2014	$12,038	$11,113	$925*** ($212)	$13,266	$12,892	$374* ($169)	$552†† ($208)
Benefits due in 2015	$12,033	$10,928	$1,105*** ($229)	$13,322	$12,755	$567** ($180)	$538† ($254)
Benefits due in 2016	$11,684	$10,510	$1,174*** ($237)	$12,847	$12,334	$513** ($185)	$661†† ($227)
Benefits due in 2017	$10,944	$9,891	$1,053*** ($244)	$12,078	$11,834	$245 ($207)	$808††† ($229)
Benefits due in 2018	$10,483	$9,619	$864*** ($256)	$11,609	$11,545	$63 ($268)	$800†† ($267)
Benefits due in 2019	$10,244	$9,578	$665** ($266)	$11,546	$11,391	$156 ($304)	$510 ($300)

Source: Analysis of SSA administrative records (from the MEF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: See Chapter 2 of the Final Evaluation Report for variable definitions. Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T21+T22 = 2,045 , Black C2 = 1,248, White T21+T22 = 4,306 , White C2 = 2,612.

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

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-5. Estimated Impacts on Any Offset Use of the Offset Compared to Current Law (T21 Vs. C2) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and WIC (T21) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and WIC (T21) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Any offset use in 2012	6.1%	0.0%	6.1%*** (0.8%)	5.7%	0.0%	5.7%*** (0.6%)	0.4% (0.9%)
Any offset use in 2013	7.6%	0.0%	7.6%*** (1.0%)	7.8%	0.0%	7.8%*** (0.6%)	-0.2% (1.3%)
Any offset use in 2014	8.4%	0.0%	8.4%*** (1.1%)	8.3%	0.0%	8.3%*** (0.7%)	0.0% (1.1%)
Any offset use in 2015	9.7%	0.0%	9.7%*** (1.2%)	10.0%	0.0%	10.0%*** (0.8%)	-0.4% (1.4%)
Any offset use in 2016	9.4%	0.0%	9.4%*** (1.1%)	9.5%	0.0%	9.5%*** (0.8%)	- 0.1% (1.3%)
Any offset use in 2017	10.1%	0.0%	10.1%*** (1.2%)	8.1%	0.0%	8.1%*** (0.7%)	2.0% (1.3%)
Any offset use in 2018	5.7%	0.0%	5.7%*** (1.0%)	2.8%	0.0%	2.8%*** (0.3%)	2.9%†† (1.1%)
Any offset use in 2019	2.8%	0.0%	2.8%*** (0.6%)	1.7%	0.0%	1.7%*** (0.3%)	1.1% (0.7%)

Source: Analysis of SSA administrative records (from the MEF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T21 = 1,253, Black C2 = 1,248, White T21 = 2,621, White C2 = 2,612.

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

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-6. Estimated Impacts on Any Offset Use of the Offset Plus WIC or EWIC Compared to Current Law (T22 Vs. C2) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and EWIC (T22) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and EWIC (T22) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Any offset use in 2012	5.8%	0.0%	5.8%*** (1.0%)	5.7%	0.0%	5.7%*** (0.7%)	0.1% (1.1%)
Any offset use in 2013	7.7%	0.0%	7.7%*** (1.3%)	8.2%	0.0%	8.2%*** (0.8%)	-0.5% (1.4%)
Any offset use in 2014	8.2%	0.0%	8.2%*** (1.1%)	9.1%	0.0%	9.1%*** (0.8%)	-0.8% (1.3%)
Any offset use in 2015	10.1%	0.0%	10.1%*** (2.1%)	9.4%	0.0%	9.4%*** (0.8%)	0.7% (2.3%)
Any offset use in 2016	9.9%	0.0%	9.9%*** (1.3%)	9.0%	0.0%	9.0%*** (0.8%)	0.8% (1.5%)
Any offset use in 2017	8.9%	0.0%	8.9%*** (1.1%)	7.1%	0.0%	7.1%*** (0.8%)	1.8% (1.3%)
Any offset use in 2018	4.8%	0.0%	4.8%*** (0.9%)	3.7%	0.0%	3.7%*** (0.5%)	1.1% (1.0%)
Any offset use in 2019	3.3%	0.0%	3.3%*** (0.7%)	1.7%	0.0%	1.7%*** (0.4%)	1.6%† (0.8%)

Source: Analysis of SSA administrative records (from the MEF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T22 = 792 , Black C2 = 1,248, White T22 = 1,694 , White C2 = 2,612.

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

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-7. Estimated Impacts on Any Offset Use of the Offset Compared to Offset Plus WIC or EWIC (T21 Vs. T22) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and EWIC (T22) (1)	Average Outcome with Offset and WIC (T21) (2)	Impact Estimate (3)	Average Outcome with Offset and EWIC (T22) (4)	Average Outcome with Offset and WIC (T21) (5)	Impact Estimate (6)	
Any offset use in 2012	5.8%	6.1%	(0.3%) (1.0%)	5.7%	5.7%	0.0% (0.9%)	0.3% (1.4%)
Any offset use in 2013	7.7%	7.6%	0.1% (0.8%)	8.2%	7.8%	0.4% (1.2%)	0.3% (1.7%)
Any offset use in 2014	8.2%	8.4%	(0.1%) (0.7%)	9.1%	8.3%	0.7% (0.9%)	0.9% (1.0%)
Any offset use in 2015	10.1%	9.7%	0.4% (1.6%)	9.4%	10.0%	(0.6%) (0.8%)	1.0% (1.7%)
Any offset use in 2016	9.9%	9.4%	0.4% (0.9%)	9.0%	9.5%	(0.5%) (0.8%)	0.9% (1.3%)
Any offset use in 2017	8.9%	10.1%	(1.2%) (1.1%)	7.1%	8.1%	(1.1%) (0.8%)	0.2% (1.4%)
Any offset use in 2018	4.8%	5.7%	(0.9%) (1.1%)	3.7%	2.8%	0.9%* (0.4%)	1.8% (1.2%)
Any offset use in 2019	3.3%	2.8%	0.5% (0.8%)	1.7%	1.7%	(0.0%) (0.5%)	0.5% (1.0%)

Source: Analysis of SSA administrative records (from the MEF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T21 = 1,253, Black T22 = 792, White T21 = 2,621, White T22 = 1,694.

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

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-8. Estimated Impacts on Any Offset Use of the Combined Offset Compared to Current Law (T21+T22 Vs. C2) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and WIC and EWIC (T22 + T21) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and WIC and EWIC (T22 + T21) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Any offset use in 2012	6.0%	0.0%	6.0%*** (0.7%)	5.7%	0.0%	5.7%*** (0.4%)	0.3% (0.6%)
Any offset use in 2013	7.7%	0.0%	7.7%*** (1.0%)	8.0%	0.0%	8.0%*** (0.5%)	0.3% (1.0%)
Any offset use in 2014	8.3%	0.0%	8.3%*** (1.0%)	8.6%	0.0%	8.6%*** (0.5%)	0.3% (0.8%)
Any offset use in 2015	9.8%	0.0%	9.8%*** (1.4%)	9.8%	0.0%	9.8%*** (0.6%)	0.1% (1.6%)
Any offset use in 2016	9.6%	0.0%	9.6%*** (1.1%)	9.4%	0.0%	9.4%*** (0.7%)	0.2% (1.2%)
Any offset use in 2017	9.6%	0.0%	9.6%*** (0.8%)	7.7%	0.0%	7.7%*** (0.6%)	1.9%† (1.0%)
Any offset use in 2018	5.3%	0.0%	5.3%*** (0.8%)	3.1%	0.0%	3.1%*** (0.3%)	2.2%†† (0.8%)
Any offset use in 2019	3.0%	0.0%	3.0%*** (0.5%)	1.7%	0.0%	1.7%*** (0.2%)	1.3%† (0.6%)

Source: Analysis of SSA administrative records (from the MEF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T21+T22 = 2,045, Black C2 = 1,248, White T21+T22 = 4,306, White C2 = 2,612.

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

†/††/††† Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-9. Estimated Impacts of the Offset Compared to Current Law (T21 Vs. C2) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and WIC (T21) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and WIC (T21) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Any benefits due in 2013	94.7%	91.5%	3.3%** (1.1%)	95.0%	93.1%	1.9%** (0.7%)	1.4% (1.3%)
Any benefits due in 2014	93.5%	88.5%	5.0%*** (1.3%)	93.1%	89.9%	3.3%*** (0.9%)	1.7% (1.6%)
Any benefits due in 2015	92.2%	85.2%	6.9%*** (1.4%)	91.9%	87.7%	4.3%*** (1.0%)	2.7% (1.7%)
Any benefits due in 2016	90.1%	82.7%	7.4%*** (1.5%)	89.9%	84.8%	5.1%*** (1.0%)	2.3% (1.8%)
Any benefits due in 2017	86.9%	79.6%	7.3%*** (2.0%)	86.8%	82.5%	4.4%*** (1.1%)	2.9% (2.0%)
Any benefits due in 2018	82.2%	77.0%	5.1%** (1.8%)	82.3%	80.5%	1.8% (1.2%)	3.3% (2.2%)
Any benefits due in 2019	78.1%	74.8%	3.3% (1.9%)	79.0%	78.5%	0.5% (1.4%)	2.7% (2.3%)

Source: Analysis of SSA administrative records (from the DAF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T21 = 1,253, Black C2 = 1,248, White T21 = 2,621, White C2 = 2,612.

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

†/††/†††Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-10. Estimated Impacts of the Offset Plus WIC or EWIC Compared to Current Law (T22 Vs. C2) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and EWIC (T22) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and EWIC (T22) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Any benefits due 2013	95.3%	91.5%	3.8%** (1.2%)	94.8%	93.1%	1.7%* (0.8%)	2.1% (1.5%)
Any benefits due 2014	94.6%	88.5%	6.1%*** (1.3%)	93.4%	89.9%	3.5%*** (1.0%)	2.5% (1.7%)
Any benefits due 2015	92.7%	85.2%	7.5%*** (1.5%)	91.9%	87.7%	4.2%*** (1.0%)	3.3% (1.8%)
Any benefits due 2016	90.0%	82.7%	7.4%*** (1.7%)	89.4%	84.8%	4.6%*** (1.2%)	2.8% (2.0%)
Any benefits due 2017	87.3%	79.6%	7.7%*** (1.8%)	86.5%	82.5%	4.0%** (1.3%)	3.7% (2.2%)
Any benefits due 2018	82.6%	77.0%	5.5%** (2.0%)	81.7%	80.5%	1.2% (1.4%)	4.4% (2.4%)
Any benefits due 2019	79.4%	74.8%	4.5%* (2.1%)	79.6%	78.5%	1.1% (1.4%)	3.4% (2.5%)

Source: Analysis of SSA administrative records (from the DAF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T21 = 792, Black C2 = 1,248, White T21 = 1,694, White C2 = 2,612.

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

†/††/†††Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-11. Estimated Impacts of the Offset Compared to Offset Plus WIC or EWIC (T21 Vs. T22) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and EWIC (T22) (1)	Average Outcome with Offset and WIC (T21) (2)	Impact Estimate (3)	Average Outcome with Offset and EWIC (T22) (4)	Average Outcome with Offset and WIC (T21) (5)	Impact Estimate (6)	
Any benefits due 2013	95.3%	94.7%	0.6% (0.6%)	94.8%	95.0%	(0.2%) (0.6%)	0.7% (0.9%)
Any benefits due 2014	94.6%	93.5%	1.1% (0.8%)	93.4%	93.1%	0.3% (0.4%)	0.8% (1.0%)
Any benefits due 2015	92.7%	92.2%	0.5% (1.2%)	91.9%	91.9%	(0.1%) (0.8%)	0.6% (1.3%)
Any benefits due 2016	90.0%	90.1%	(0.0%) (1.6%)	89.4%	89.9%	(0.5%) (1.0%)	0.5% (1.9%)
Any benefits due 2017	87.3%	86.9%	0.4% (1.2%)	86.5%	86.8%	(0.3%) (1.2%)	0.7% (1.8%)
Any benefits due 2018	82.6%	82.2%	0.4% (1.4%)	81.7%	82.3%	(0.7%) (1.5%)	1.1% (2.2%)
Any benefits due 2019	79.4%	78.1%	1.3% (1.4%)	79.6%	79.0%	0.6% (1.6%)	0.7% (2.2%)

Source: Analysis of SSA administrative records (from the DAF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T21 = 1,253, Black T22 = 792, White T21 = 2,621, White T22 = 1,694.

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

†/††/†††Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.

Exhibit B-12. Estimated Impacts of the Combined Offset Compared to Current Law (T21+T22 Vs. C2) for Subgroups Defined by Black versus White

Outcome	Black			White			Estimated Difference in Impact (7)
	Average Outcome with Offset and WIC and EWIC (T22 + T21) (1)	Average Outcome under Current Law (C2) (2)	Impact Estimate (3)	Average Outcome with Offset and WIC and EWIC (T22 + T21) (4)	Average Outcome under Current Law (C2) (5)	Impact Estimate (6)	
Any benefits due 2013	95.0%	91.5%	3.5%*** (1.0%)	94.9%	93.1%	1.8%** (0.7%)	1.7%† (0.8%)
Any benefits due 2014	93.9%	88.5%	5.4%*** (1.1%)	93.2%	89.9%	3.4%*** (0.8%)	2.0%†† (0.9%)
Any benefits due 2015	92.4%	85.2%	7.1%*** (1.2%)	91.9%	87.7%	4.2%*** (0.9%)	2.9%†† (1.0%)
Any benefits due 2016	90.0%	82.7%	7.4%*** (1.4%)	89.7%	84.8%	4.9%*** (0.9%)	2.5%†† (0.8%)
Any benefits due 2017	87.0%	79.6%	7.4%*** (1.7%)	86.7%	82.5%	4.2%*** (1.0%)	3.2%†† (1.3%)
Any benefits due 2018	82.3%	77.0%	5.3%*** (1.6%)	82.1%	80.5%	1.6% (1.1%)	3.7%†† (1.2%)
Any benefits due 2019	78.6%	74.8%	3.8%* (1.7%)	79.3%	78.5%	0.8% (1.2%)	3.0%† (1.3%)

Source: Analysis of SSA administrative records (from the DAF, BODS, MBR, and SSR), with covariates from Stage 2 baseline survey and baseline SSA administrative data used in impact analysis regression equations. Self-reported race and ethnicity data come from the Stage 2 baseline survey.

Notes: Weights reflecting sample selection are used to ensure that the BOND subjects who met analysis criteria are representative of volunteers for offset participation in the nation. Standard errors are in parentheses. Means and impact estimates are regression-adjusted for baseline characteristics.

Unweighted sample sizes: Black T21+T22 = 2,045, Black C2 = 1,248, White T21+T22 = 4,306, White C2 = 2,612.

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

†/††/†††Difference in impact estimates is significantly different from zero at the .10/.05/.01 levels, respectively, using an F-test.