

Can changes in disability insurance work incentives influence beneficiary employment? Evidence from the Promoting Opportunity Demonstration*

Michael Levere[^], David Wittenburg[&], and John Jones⁺

April 2024

We study how disability beneficiary work behavior responds to a rule change that replaces a cash cliff—a threshold above which benefits reduce to zero—with a benefit offset ramp. Under existing rules, beneficiaries can test work though risk losing benefits with prolonged earnings that exceed that key threshold. With the offset ramp, benefits could adjust each month based on the previous month's earnings. Using a randomized controlled trial with over 10,000 Social Security Disability Insurance beneficiaries who voluntarily enrolled in the demonstration, we find precisely estimated null effects on earnings, income, and benefit amounts. An analysis of mechanisms indicates that administrative burden, the limited size of the incentive, and individual and systemic barriers to employment for people with disabilities likely contributed to the limited impacts.

[^] Colgate University and Mathematica

[&] Mathematica

⁺ Social Security Administration

* The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA). The findings and conclusions are solely those of the author(s) and do not represent the opinions or policy of SSA or any agency of the Federal Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation, or favoring by the United States Government or any agency thereof.

I. Introduction

There is strong policy interest in assessing whether Social Security Disability Insurance (SSDI) program rules affect beneficiary work activity. SSDI is the largest cash transfer program for former workers with disabilities. Prior evidence indicates SSDI programmatic rules are complex to administer and potentially discourage beneficiaries from working (Gelber et al. 2017; Maestas et al. 2013). The SSDI program includes provisions that allow beneficiaries to test work without risking loss of benefits, such as a Trial Work Period (TWP). If earnings are sustained above a key threshold for an extended period (the substantial gainful activity [SGA] amount), SSDI beneficiaries face the “cash cliff,” where they risk losing their entire SSDI benefit. Despite a seemingly strong incentive to avoid work above this threshold, there is limited evidence in the United States that this cash cliff provision substantially affects work activity (Schimmel et al. 2011; Weathers and Hemmeter 2011; Gubits et al. 2018). Outside the United States, evidence from Austria indicates people bunch earnings just below another key earnings threshold (Ruh and Staubli 2019).

We explore the impact of modifying work incentive rules for SSDI beneficiaries by replacing the cash cliff with a benefit offset from the Promoting Opportunity Demonstration (POD). This “benefit offset” reduces benefits by \$1 for every \$2 in earnings above a certain amount. In addition, the rule changes also simplified other provisions of current rules to allow the benefit offset to apply immediately rather than have an extended period for beneficiaries to test work activity before work affects benefits. SSA applied the benefit offset monthly, thus requiring beneficiaries to report their earnings each month.¹ Beneficiaries with earnings

¹ Beneficiaries with the same earnings each month did not need to report their earnings every month, as the offset adjustment was based on the most recent amount of earnings reported. However, beneficiaries were still encouraged

sufficiently high to use the offset² might work more when the uncertainty of a cash cliff is removed. Yet to maximize income under these rules, they must both understand the rules and sustain their work effort.

We use a randomized controlled trial with more than 10,000 self-selected SSDI beneficiaries to explore the impact of the POD rule changes two years after enrollment. Treatment group members were subject to the benefit offset, while control group members faced current rules (including the cash cliff).³ Because of the randomized design, any differences in outcomes between the treatment and control groups represent the causal effect of the rule changes. We examine impacts on outcomes related to benefit amounts, employment, and earnings from SSA administrative data. We use POD survey data to capture information not available in the administrative records to provide additional information on impacts related to employment activity (e.g., work search), income, health, and receipt of other program benefits. We also draw on data collected through the POD implementation in supplemental analyses to help understand the mechanism behind the impacts (or lack thereof).

These rule changes had limited impacts over the two years after enrollment. Average earnings, SSDI benefits, and income are essentially identical between the treatment and control groups. The estimated impacts are also sufficiently precise to rule out substantive changes. We did find one statistically significant impact on the percentage of people who had “substantive

to report earnings each month, particularly because earnings of people with disabilities can be particularly volatile (Jolly and Wagner 2023).

² Throughout the rest of the paper, we refer to beneficiaries with earnings above the offset threshold as using the offset to simplify the language.

³ The treatment group technically consisted of two separate treatment groups. One treatment group faced termination if they had 12 consecutive months of earnings sufficiently high to lead benefits to fall to zero. The other treatment group did not face termination due to excess earnings. However, work activity and use of the benefit offset were nearly identical across the two treatment groups. Further, fewer than 1 percent of the termination-possible treatment group faced the termination provision. Therefore, we pool together these two treatment groups throughout this paper and conduct analysis as if there was only a single treatment group.

earnings” (measured as earning above roughly \$15,000). In this case, treatment group members were about 1 percentage point more likely, or 10 percent relative to the control group mean, to have substantive earnings. Impacts on other outcomes related to employment activity, health, and receipt of other program benefits were also mostly limited. In conducting subgroup analyses, we found that the rule changes had positive impacts on earnings and income for beneficiaries not employed at baseline. For those employed at baseline, the results are not statistically significant for most outcomes. However, these impacts were not significantly different across these two subgroups.

We then explore the reasons why impacts on earnings and benefit amounts were limited despite the substantial change in rules and thus incentive to work, highlighting three potential mechanisms that contributed to limited impacts. The first mechanism is administrative burden, which can be any sort of hassle or challenge people face when interacting with a government program (Herd and Moynihan 2019). These can include learning, compliance, and psychological costs. Though administrative burden decreased in many important ways with the demonstration, three critical types of learning costs remained. First, treatment group members had limited understanding of the program rules. Only 46 percent of treatment group members correctly understood that the benefit offset reduced their monthly benefits for earnings above a key threshold, the fundamental premise of the demonstration. Second, many participants experienced overpayments that may have exacerbated issues around understanding how work activity related to adjustments in benefits. More than one in five treatment group members experienced an overpayment (or nearly three in four among the roughly one-third who had their benefits adjusted), wherein the beneficiary then would owe back benefits received during a period where

benefits should instead have been reduced because of work activity.⁴ Third, the limited timespan for the demonstration of only two years may have meant that people in both the control and treatment group did not have long enough exposure to the work incentives to adjust their employment intentions. For example, for those in the control group, it is possible that a longer time horizon might be important as people finish their TWP and, hence, are subject to the cash cliff. For the treatment group, it could take time, especially for those who were not working at the start of the demonstration, to learn the new rules. . Throughout the demonstration, particularly at the early stages of the demonstration, treatment group members learned about new processes like regularly reporting earnings timely, highlighting both that the time frame was short and that the burdens of overpayments and limited knowledge may have ultimately improved if the demonstration ran longer. Though the demonstration clearly reduced burdens for beneficiaries along multiple dimensions, these three factors indicate that enough burdens may still have remained such that it limited impacts.

A second issue was the size of the incentive itself. For some, the new offset rules represented a disincentive to work. People whose benefits were suspended at the time of enrollment would now have less incentive to work both because of an income effect (they now get additional income in the form of a benefit payment) and a substitution effect (each \$1 of earnings now leads to only \$0.50 of additional take-home pay). Though this may be a relatively small group, this group may already have higher work activity and contribute heavily to the overall effects. However, we analyze subgroup effects among various groups that might represent those where the offset theoretically provided a disincentive for work and do not find

⁴ Importantly, the overpayments for those subject to the new rules represents a very different type of burden than these beneficiaries likely otherwise would have experienced under current SSDI rules. Under current rules, overpayments are less frequent, but when they occur are for large sums of money (Hoffman et al. 2019). Under the demonstration rules, overpayments were much more frequent but for a much smaller dollar amount.

such effects (though caution that these are very small subgroups). Additionally for people with earnings just below the cash cliff amount, their benefit would decline relative to current SSDI rules. Such people would have lower work incentives because of the substitution effect (a 50 percent higher marginal tax rate because of the \$1 for \$2 offset) compared to before the demonstration, though would also have more incentive to work because of an income effect (with a smaller benefit amount). About one in four offset users had earnings amounts that led them to fall in this range during the demonstration, suggesting this may play an important role.

A third contributing factor was systemic and individual barriers to employment. At enrollment, 90 percent agreed it was difficult to work because of a physical or mental condition. Employment activity was limited among control group members: more than half of people had no earnings over the two-year study period, while only 10 percent of people had earnings above about \$15,000 per year. Together with evidence around limited work activity for people with disabilities generally (e.g., Livermore and Honeycutt 2015), potentially because of discrimination (e.g., Bellemare et al. 2023), this suggests that even if people wanted to work to take advantage of the rules, they may not have been able to.

Our paper adds to existing literature on the complexities of enhancing work activity among SSDI beneficiaries. Studies like Maestas et al. (2013), French and Song (2014), Gelber et al. (2017), and Moore (2015) indicates limited work activity among these beneficiaries, attributed partly to SSDI participation itself. These studies exploit exogenous factors related to program administration (such as variation in judge stringency) to understand the casual effects of SSDI receipt. Nichols et al. (2021) further elucidate this by discussing the modest impact of SSA's work incentives, potentially due to SSDI beneficiaries' limited work capacity. Our analysis aligns with these findings, underscoring the restricted work activity in this group.

Our findings also provide insights into the importance of administrative burden in government benefits programs. Typically, the literature highlights ways that administrative burden affects program enrollment. For example, Deshpande and Li (2019) show how field office closures, which make it harder for people to apply for benefits, reduce applications to Supplemental Security Income (SSI). Homonoff and Somerville (2021) show that the recertification process for the Supplemental Nutrition and Assistance Program leads fewer people to participate. Our findings contribute to a relatively new literature on redemption costs (Barnes 2021), which highlight ways that administrative burden can prevent people from effectively using a program.

This paper proceeds as follows. Section II offers more details on the institutional context related to the SSDI program and on how exactly POD changed the program rules. We describe in Section III the data and methods that we use. Section IV presents the primary results. Section V explores the mechanisms, highlighting potential reasons that impacts may be limited. Finally, Section VI concludes.

II. Institutional Context

A. Overview of current SSDI rules

To qualify for SSDI benefits, an individual must be unable to engage in work that constitutes SGA. Earnings above the SGA amount are typically considered evidence that the beneficiary does not have a work-limiting impairment and therefore is ineligible to receive SSDI benefits.

During a non-consecutive 12-month period, which includes a 9-month Trial Work Period (TWP) and a 3-month grace period, beneficiaries receive a full SSDI benefit check regardless of how much they earn. After this, SSA generally suspends beneficiaries' full cash benefits if their

countable earnings reach or exceed the SGA level (the cash cliff). TWP months are counted within a 5-year rolling window. After completing the TWP, a beneficiary immediately enters the Extended Period of Eligibility (EPE). The EPE is a 36-month re-entitlement period, during which benefits are suspended for months in which earnings exceed the SGA amount (with the exception of the three-month grace period) and full benefits are paid for months in which earnings fall below the SGA level (i.e., they are not adjusted for any earnings below that amount).⁵

SSA terminates benefits if earnings exceed the SGA level after the re-entitlement period (the EPE) ends and the beneficiary has used all grace-period months. Otherwise, benefit payments continue in full. If benefits are terminated due to SGA, beneficiaries can seek expedited reinstatement of benefits at any point during the 60 months following their notification of benefit termination by SSA. Substantive earnings activity among SSDI beneficiaries can also affect their Medicare eligibility after an extended period.⁶

SSA has a long-standing interest in understanding options to promote employment among SSDI beneficiaries. For example, SSA tested a different benefit offset that also eliminated the cash cliff as part of the Benefit Offset National Demonstration (BOND). The changes in program rules under BOND applied at a higher level of earnings and only after the beneficiary had completed the TWP and the grace period. BOND had minimal effects on earnings and led to increases in SSDI benefit payments (Gubits et al. 2018). The design of POD, which was Congressionally mandated, drew on lessons from that previous experience. For example, POD

⁵ In making this SGA determination, SSA uses an adjusted measure of earnings that deducts SSA-approved impairment related work expenses and other noncountable income such as sick pay, vacation pay, and subsidies.

⁶ SSDI beneficiaries become eligible for Medicare Part A benefits (and can pay a monthly premium to receive Medicare Part B benefits) 24 months after SSDI eligibility. Beneficiaries with cash benefits terminated based on the performance of SGA generally lose their Medicare benefits 93 months after completion of the TWP.

addresses the perception that BOND rules were complex by using a simplified set of administrative adjustments in implementing the offset, as discussed next. This attempt is important because it could reduce the administrative burden to treatment group members in understanding the implications of work rules. However, POD also included other provisions that could increase the administrative burden to treatment group members, especially monthly earnings reporting.

B. The Promoting Opportunity Demonstration

The rules for POD included a benefit offset ramp and modified other current rule provisions. This offset reduced benefits by \$1 for every \$2 of monthly earnings higher than the POD threshold (equal to the TWP amount, which was \$940 in 2021).⁷ The benefit offset applied immediately to monthly earnings, eliminating the TWP, grace period, and EPE. This change made it more straightforward to describe how earnings affected benefit adjustments, and thus likely reduces administrative burden. However, it also required that treatment group members report earnings monthly for their benefits to be adjusted promptly, which introduced a different administrative burden.⁸

SSA created a separate infrastructure to implement this benefit offset and support beneficiaries through the process. SSA funded an implementation team that created a system to collect earnings information (e.g., pay stubs) from treatment group members. This team then sent

⁷ POD also includes special provisions for beneficiaries who have Impairment-Related Work Expenses (IRWE). For such beneficiaries, the threshold is defined as the greater of the TWP amount and a beneficiary's IRWE (up to a maximum of the SGA amount). However, such beneficiaries are rare, with fewer than one percent of POD participants having a threshold higher than the TWP amount.

⁸ Importantly, though, under current rules beneficiaries need to report whenever they start or stop work, if duties, hours, or pay changes, or there are impairment-related work expenses. Thus, this may not represent a substantial change to the reporting process for beneficiaries who work.

the information to SSA, who recorded those earnings as received and, in most cases, used an automated system to process benefit adjustments. At the end of the year, SSA checked if an adjustment was necessary for earnings not reported.⁹

Recruitment efforts included both direct and indirect outreach to inform beneficiaries of the rule changes. Direct outreach included informational mailings sent to potentially eligible beneficiaries, while indirect outreach included efforts to raise awareness through a website and communications with community organizations that serve SSDI beneficiaries. Benefits counselors also offered support throughout the demonstration. Counselors contacted treatment group members upon enrollment to provide information about the new POD rules. The counselors also provided referrals and more in-depth work incentive counseling throughout the demonstration based on the treatment group members' needs; the latter type of counseling is also available under current program rules.

10,070 working-age SSDI beneficiaries in the eight POD states (Alabama, California, Connecticut, Maryland, Michigan, Nebraska, Texas, and Vermont) volunteered for POD and provided written informed consent between January 2018 and January 2019.¹⁰ Once enrolled, people were eligible to use the benefit offset until June 2021, when the demonstration ended, at which point they returned to current-law rules. A key feature of the demonstration authority was

⁹ Specifically, SSA used the POD automated data system to sum each treatment group member's monthly earnings reports submitted across all months in the year and compared them with the total annual gross earnings from Internal Revenue Service records. This process allowed SSA to determine the SSDI benefits that should have been paid to each POD treatment group member during the previous calendar year and compare it to the actual amount of SSDI benefits paid.

¹⁰ This section represents a synthesis of findings from Wittenburg et al. (2022).

that participation was voluntary,¹¹ and, hence, participants could withdraw from the demonstration at any time.

The COVID-19 pandemic overlapped with part of the demonstration period, during which SSA made several adjustments for SSDI beneficiaries generally and POD treatment group members specifically to maintain continuity of services. These changes allowed the demonstration (and current SSDI services) to continue despite broad disruptions from the pandemic. First, SSA severely limited in-person services at its local field offices and reprioritized workloads at the start of the pandemic in March 2020. Second, SSA added protections to help avoid disruptions to benefit payments for all SSDI beneficiaries. Finally, counselors called all treatment group members in the early phase of the pandemic to offer support, connect them to area resources, and inquire about changes in their employment status. Importantly, we do not find substantive differences in impacts over the pandemic period.¹²

Beneficiaries who enrolled in the demonstration differed from the general SSDI beneficiary population, particularly in their interest in work. For example, 15 percent of POD enrollees had earnings at or above the SGA amount since 2014, which was about 2.5 times the rate for non-volunteers.¹³ This group, which would be approaching benefit suspense depending on their completion of the TWP, is particularly notable given that POD represents a work disincentive for anyone who has benefits already in suspense at enrollment. POD enrollees also differed from beneficiaries who did not volunteer along other characteristics, though many of

¹¹ The 10,070 beneficiaries who enrolled represented 2.4 percent of those solicited to sign up through direct outreach.

¹² Specifically, our findings are similar for 2019 alone, 2020 alone, or the average across both years (the latter corresponding to the primary findings presented in this paper). Average earnings in 2020 were roughly similar to earnings in 2019 for both the treatment and control groups despite the pandemic.

¹³ For details on other characteristics, see Exhibit D.8 in Hock et al. (2020).

these differences may also stem from enrollees' stronger connection to work. For example, POD enrollees were younger and were more likely to have a mental disorder than non-volunteers. The implication of these differences is that the evaluation findings are specific to the sample of POD enrollees. In other words, our findings represent the impacts on the beneficiary population who decided to enroll and do not necessarily generalize to a nationally representative population of SSDI beneficiaries. Nonetheless, because those who enrolled have a strong interest in working, we may have been more likely to find a significant impact of the demonstration than among a randomly sampled group of SSDI beneficiaries.

Importantly, the theoretical predictions of the offset on work behavior varied depending on beneficiary earnings. For example, the POD rules were financially favorable for beneficiaries consistently earning above the current SGA amount, particularly if they had completed the TWP and grace period. Under POD rules, these beneficiaries received cash benefits reduced by half of the difference between their monthly earnings level and the POD threshold, rather than \$0 if they faced the cash cliff under current rules. However, this leads them to have less incentive to work: theoretically, the re-introduced benefit would incentivize leisure through an income effect, while the phase-out and 50 percent marginal tax rate would incentivize leisure through a substitution effect. In contrast, some beneficiaries, such as those still in the TWP, would fare worse under POD rules relative to current rules. POD resulted in lower total income for treatment group members if they had not completed the TWP and grace period or if they earned between the TWP threshold and the SGA amount. Beneficiaries in the former group would not yet lose benefits under current rules regardless of earnings, whereas the benefit offset applied immediately. Beneficiaries in the latter group would be eligible for full benefits under current law, but instead would have their benefits partially offset. POD therefore might have ambiguous

impacts on benefits and earnings, particularly depending on a beneficiary's recent connection to work.

III. Data

Our data included information from SSA program records, earnings reported to the IRS, the POD Implementation Data System, and three surveys. Together, these data enabled us to examine offset use, assess understanding of earnings rules, assess experiences with improper payments, and to estimate program impacts on employment, benefits, and other outcomes.

Data on SSA program participation and earnings comes from several SSA data sources, which we used to construct all the primary outcomes for the impact analysis. We use the Master Beneficiary Record and the Supplemental Security Record to track monthly SSDI and Supplemental Security Income (SSI) program participation, respectively. We construct outcomes corresponding to benefit amounts due in the first 24 months after enrolling in the demonstration. Our employment and earnings measures from the Master Earnings File represented average annual earnings reported to the IRS. The annual earnings data covered 2019 and 2020, which encompassed the two calendar years after the year of enrollment.¹⁴ We also construct measures of average annual income, which include earnings plus SSDI and SSI benefit amounts due in 2019 and 2020.¹⁵ Our analysis of improper payments is based on monthly snapshots from the Master Beneficiary Record. To identify improper payments, we look at whether the cash benefit

¹⁴ About 2 percent of participants were enrolled and randomly assigned in January 2019 (Hock et al. 2020). However, because these beneficiaries had to submit their enrollment materials before December 31, 2018, outcomes measured in calendar years 2019 and 2020 are still a good proxy for their experience in the first two years after enrollment. To maintain consistency, we essentially treated December 2018 as the month of enrollment for beneficiaries who enrolled in January 2019.

¹⁵ This is of course only a partial measure of income, but captures what we are reliably able to measure from SSA administrative records.

due at the time the payment was made is different from the cash benefit due for that month as of June 2021.¹⁶

We also use information from the POD Implementation Data System and three surveys to assess additional outcomes and probe the mechanisms driving our findings. Data on benefit offset usage comes from the Implementation Data System,¹⁷ which we report on for the first 24 months after enrolling in the demonstration. Our survey data efforts included a baseline survey and two follow-up surveys. Beneficiaries had to complete the baseline survey to enroll in POD. The two follow-up surveys included content on follow-up activities one and two years after random assignment. Both surveys captured information about enrollees' employment, understanding of program rules, attitudes about work, income, health and functional status, and health insurance. More than 80 percent of surveyed beneficiaries completed the follow-up surveys (84 percent for the first follow-up survey and 83 percent for the second follow-up survey). The first follow-up survey included a random sample of half of POD enrollees,¹⁸ while the second follow-up survey included the full sample of POD enrollees. Response rates did not differ for the treatment and control groups.

IV. Methods

We used a randomized controlled trial to test the impact of the modified rules. The 10,070 beneficiaries who enrolled in the demonstration were randomly assigned to either a

¹⁶ For more details on how we measure improper payments, see Appendix E and Section VI.3 in Wittenburg et al. (2022).

¹⁷ The POD implementation data system, which the implementation team maintained, included information on POD related services, such as the provision of work incentive counseling, collection and submission of earnings information to SSA, offset use, and transition back to program rules at the end of subjects' POD participation period.

¹⁸ We designed the random sampling procedure to guarantee that the characteristics of those who were selected to participate in the survey closely resembled the characteristics of all POD enrollees. The random selection plus similar characteristics of the survey sample means that the estimates from the survey data are representative of all POD enrollees.

treatment group (6,700 beneficiaries) or control group (3,370 beneficiaries). Random assignment means that control group members should represent a valid counterfactual for what outcomes of treatment group members might have looked like if not for their participation in POD.

We did not find any substantive differences across any major demographic, impairment or work history categories between control and treatment groups (Table 1).

Equation (1) shows our primary estimating equation. We estimated impacts using an ordinary least-squares model with heteroskedasticity-robust standard errors.

(1)

$$y_i = \alpha + \beta T_i + \delta X_i + \varepsilon_i$$

The coefficient of interest is β , which measures the adjusted difference in means between the treatment and control groups. Because of the randomized design, β therefore represents the causal impact of POD. We control for several individual characteristics in X_i , including those used to stratify random assignment¹⁹ and an array of baseline characteristics, such as sex (collected in administrative records), race (collected in the baseline survey), recent work activity (also in the baseline survey), and three characteristics that exhibited chance statistical differences between the groups at baseline.²⁰ Our results are robust to the exclusion of these individual characteristics. Except for the characteristics used to stratify random assignment, including these

¹⁹ We stratified random assignment by state. Within each state, we also stratified in the following ways: first, if someone had one of three rare diagnoses (neoplasms, injuries, or severe visual impairments), we only stratified by the primary diagnosis. For everyone with a different diagnosis, we stratified by state, age groups (either ages 20 to 34, 35 to 44, or 45 and older), SSDI benefit duration (1 to 18 months, 19 to 36 months, and 36 months or more), and whether they reported earning over \$1,000 per month in the baseline survey at enrollment.

²⁰ These three characteristics are all measured in the baseline survey. They include the extent to which people agree that it is difficult to work because of fear of losing disability cash benefits; it is difficult to work because of fear of losing health insurance; and it will be difficult to receive SSDI in the future if one works.

characteristics as controls should only improve the precision of the impact estimates but they are not needed to generate unbiased results.

All impact estimates for the POD evaluation are intent-to-treat estimates. These estimates measure the effects of being assigned to POD rules on treatment group members (relative to control group members), regardless of their post-enrollment behavior. In other words, we estimate the impacts on all POD enrollees, irrespective of whether they actively engaged with it.

Our analysis includes administrative data for all POD enrollees and, for survey outcomes, the subset of enrollees who responded to surveys. For the latter, we use weights to produce estimates that reflect the impact of POD rules on all POD enrollees. Analysis weights for outcomes from the one-year follow-up survey account for survey sampling and nonresponse,²¹ while the analysis weights for the two-year follow-up survey only account for nonresponse as the survey included all enrollees.

We pre-specified four primary outcomes as the main assessment of POD's efficacy to avoid problems related to multiple comparisons. Multiple comparisons can cause problems when many statistical tests are performed. We address this issue by pre-specifying four primary outcomes. By choosing these outcomes before conducting the analysis from among the dozens available, we reduce the likelihood of finding impacts by chance alone without significantly undermining the statistical power of the evaluation to detect true impacts. We operationalize this

²¹ The weights are the product of two terms: sampling weights and the survey nonresponse weights. The sampling weight (the first term) is determined by the probability of being sampled for that survey. Because we randomly sampled half the POD enrollees for the year-one follow-up survey, the sampling weight term in the overall weight is the same for all POD enrollees. To construct the survey nonresponse weight (the second term in the overall weight), we use a random forest algorithm. The algorithm uses observable baseline characteristics to predict the probability that each person responded to the survey. The nonresponse weight equals the inverse of the estimated response probability.

approach in the presentation of findings by placing greater emphasis on the interpretation of primary than of secondary outcomes.

The four primary outcomes include continuous measures of earnings, benefit payments, and income, as well as an indicator for having substantive earnings, or about \$15,000 in earnings (more precisely, the annualized SGA amount). Secondary outcomes include several employment-related outcomes and several health and health insurance related outcomes. For employment-related outcomes, we report on information collected from the survey (any employment in the past year, being employed or actively searching for a job, hours worked, fringe benefits), SSA program records (whether earnings were greater than \$0), and vocational rehabilitation (VR) program records (whether people applied for or received VR services). We also analyze measures of physical and mental health based on the 12-item Short Form Survey developed from the Medical Outcomes Study (Hays et al. 1995), as well as whether the beneficiary reports having any health insurance coverage.

We also conducted several subgroup impact analyses to assess whether the modified rules had differential impacts on certain types of beneficiaries. In this paper, we focus on the subgroup defined by employment status at the time of enrollment, which is self-reported in the baseline survey.²² About 23 percent of enrollees were employed at enrollment versus 77 percent who were not employed. This group is of interest given enrollees who were employed at baseline are potentially more likely than other enrollees to use the benefit offset (Gubits et al. 2018). We estimate the mean impact of assignment to the treatment group for each of the two subgroups, as well as run a statistical test for whether the mean impacts are different from each other.

²² In alternative analyses, we also consider subgroup analysis based on work expectation at the time of enrollment, level of education, age, primary impairment, or state of residence. However, we found essentially no notable differential impacts across subgroups for the primary outcomes, so to save space we do not report such findings in this paper. For the results of those analyses, please see Exhibits F.2 to F.6 in Wittenburg et al. (2022).

V. Results

A. Primary impact results

The intervention had no overall impact on earnings (Table 2). Average annual earnings in 2019 and 2020 for treatment group members was \$5,022, relative to \$4,954 for the control group. The estimated difference of \$68 represented about 1 percent of the control group mean, which implies no substantive change in outcomes. The 95 percent confidence interval covers the range of a decrease in earnings of \$323 to an increase of \$460. We can therefore rule out an effect on earnings of more than 10 percent.

Those assigned to the treatment group were 1 percentage point more likely to have substantive earnings, defined as annual earnings above approximately \$15,000 (the annualized SGA amount), and this difference was statistically significant at the 10 percent level. We expected that those who would benefit most from the modified work rules would be those who could earn above this SGA amount, the threshold at which the cash cliff applies, on a continuous basis. About 11 percent of the treatment group had such earnings, compared with 10 percent of the control group; the estimated difference of 1 percentage point represents a 10 percent increase relative to the control group mean.

We also found no significant differences between the treatment and control groups in terms of SSDI benefit amounts or income. The average annual SSDI benefit amounts in the two years after enrolling for the treatment group was \$11,870. Total income for treatment group members, which included total earnings plus SSDI benefit amounts and SSI payment amounts, was \$16,775. For both measures, the control group mean was within 2 percent of the treatment group mean, which further underscores the interpretation of no impact. Standard errors are also precisely estimated: we can rule out changes of more than 4 percent in either direction.

The lack of impacts on benefit amounts is notable given that 30 percent of treatment group members used the benefit offset, as discussed further below. The rules associated with the benefit offset would lead some treatment group members to experience increased benefit amounts without any changes in earnings behavior (such as those who had completed the TWP and had substantive earnings). However, the new rules would also decrease the benefits of other treatment group members (for example, those in the TWP). Thus, a null effect may mask heterogeneity in that there were increases in benefit amounts for some and decreases for others. As shown in Appendix Table 1, benefit amounts increased by nearly \$3,000 per year for the small group of beneficiaries whose benefits were suspended at the time of enrollment.

When we analyze heterogeneous impacts by employment status at enrollment, we find that POD had positive and significant impacts on earnings, annualized SGA amount, and income within the subgroup of beneficiaries not employed (Table 3, Panel B). The magnitude of these impact estimates was large in percentage terms relative to the control group mean. For earnings, the increase of \$298 represents a 14 percent increase for treatment group members relative to the control group mean (within the subgroup of beneficiaries not employed at baseline). For substantive earnings, the increase in prevalence of 1.2 percentage points represents a 29 percent increase relative to the control group mean within this subgroup. However, impacts for those not employed at baseline were not significantly different from impacts for those employed at baseline, indicating these results should be considered as exploratory.

B. Supplemental impact results

POD had relatively limited effects on most secondary outcomes and subgroups, though there were some modest effects on employment-related activities (Table 4).²³ Treatment group members were about 3 percentage points more likely than control group members to report that they were either employed or actively searched for a job in the year before the survey. This impact represented a 5 percent increase relative to the control group mean. We also found that treatment group members had more active engagement with VR services than control group members. Those in the treatment group were 1.3 percentage points more likely to apply for and 0.7 percentage points more likely to use VR services. Though these differences are small in magnitude, they are large relative to the control group mean (about 50 and 20 percent, respectively). In contrast, we found no difference on most outcomes related to health and health insurance. For example, about 99 percent of both treatment and control group members had health insurance coverage, and average physical and mental health scores were similar.

VI. Mechanisms

In this section, we explore potential reasons that might explain the limited impacts on primary outcomes presented in Section V. We point to three main factors: (1) administrative burden (specifically, learning costs); (2) the incentive not being big enough; and (3) individual or systemic barriers to employment for people with disabilities more generally. We present evidence related to each of these three main barriers, suggesting that each likely play a role in the limited impacts. However, because the analyses in this section are primarily descriptive (i.e., they do not rely on the randomized design of the intervention), they should be viewed as suggestive in nature.

²³ We focus here on findings in the two-year follow-up survey, which included all enrollees. Results from the one-year follow-up survey are mostly similar, and are available upon request or in Wittenburg et al. (2022).

A. *Administrative burden*

The rule changes associated with the benefit offset simplified existing rules and thus limited learning costs and reduced administrative burden in several ways. For example, the rules eliminated the TWP and grace period associated with current SSDI rules, which a beneficiary would have to go through before facing the cash cliff.²⁴ Beneficiaries in the control group (for whom these rules continued to apply) had difficulty understanding these rules: only about one-quarter knew that a TWP existed in which benefits remain unchanged regardless of earnings (Table 5, Panel B). POD also eliminated work-related continuing disability reviews.

The rules also changed the burden associated with improper payments, which occur when SSA pays beneficiaries more (or less) in SSDI benefits than they were entitled to based on work activity. Under current rules, improper payments are quite rare, but overpayments impose a substantial financial burden to beneficiaries who experience them: Hoffman et al. (2019) show that from 2010 to 2012, only about 1.9 percent of SSDI beneficiaries faced overpayments, but for those who did, the median overpayment amount was more than \$9,000 (which was then owed back to SSA). The large amount stems from the fact that monthly SSDI benefit payments are all or nothing – if the benefit is overpaid, it is by definition overpaid by the full benefit amount. By introducing a benefit offset, where beneficiaries can still be paid a partial benefit in a month, the new rules dramatically limited the size of these overpayments.

However, there were reports of other types of learning costs that might have contributed to the limited impacts. Though the new rules reduced the amount of administrative burden, the key question is whether it reduced administrative burden by *enough*, rather than if it reduced

²⁴ These provisions might be valuable work incentives for beneficiaries as they allow people to test work before benefits can become suspended.

administrative burden *at all*. Under the new benefit offset rules, learning costs remained for beneficiaries. POD treatment group members also struggled to understand the new offset rules and many struggled with filing monthly earnings on time, which led to overpayments. Below, we provide evidence on these factors.

One important challenge was that treatment group members had substantial difficulty understanding the rules. In the two-year follow-up survey, only 46 percent of treatment group members correctly understood that the benefit offset reduced their monthly benefits for earnings above a key threshold (Table 5, Panel A).²⁵ Only 34 percent correctly identified whether their benefits could be terminated if their earnings were too high.²⁶ Understanding of the rules was approximately similar in surveys conducted both one and two years after enrollment, indicating that despite opportunities to further engage with the new rules, knowledge did not improve. Perhaps not surprisingly, offset users were more likely than those who never used the benefit offset to correctly answer those questions. Nonetheless, a particular challenge in this context may relate to the fact that all POD enrollees had some exposure to current rules before enrolling in the demonstration. Thus, knowing how the *change* in the rules might influence the *change* in optimal earnings behavior might be especially difficult.

A second important challenge was that people in the treatment group who used the benefit offset faced frequent improper payments, which may have exacerbated issues around

²⁵ The precise survey question asked respondents “Under POD, do you have a TWP where your benefits remain unchanged regardless of your earnings?” Beneficiaries had the option to respond that they did not know the answer, in which case they are considered to not correctly understand the concept.

²⁶ The precise survey question asked respondents “Under the POD rules, do your benefits ever terminate if your earnings are too high?” The correct answer depended on the version of the POD rules. Beneficiaries had the option to respond that they did not know the answer, in which case they are considered to not correctly understand the concept.

understanding how work activity related to adjustments in benefits. Among those who used the offset in 2019, 86 percent had an improper payment.²⁷ About 74 percent had an overpayment, and about 40 percent had an underpayment. Because about 30 percent of treatment group members used the benefit offset (discussed below), this means more than one in five treatment group members experienced an overpayment. The high prevalence of improper payments may make it challenging for beneficiaries to correctly anticipate the way that their work activity will affect their benefits because it attenuates the connection between earnings and benefit amounts. An important contributor to these overpayments was the need to report earnings monthly in a timely fashion: as discussed next, many people faced challenges with reporting timely, though this improved as the demonstration went on. Overpayments were pervasive and not simply one-time events: offset users had an overpayment in about half (46 percent) of offset months. Among all offset users, the median total overpayment amount was \$482, and each offset user with overpayments experienced a median of 2.5 months of overpayments. Thus, though the new rules substantially limited the burden associated with overpayments by leading to much smaller overpayment amounts, a significant burden remained through the high frequency of overpayments.

Finally, the short timespan of the demonstration (two years) may have limited the potential for impacts to emerge. Specifically, Congress mandated that the entirety of the demonstration—including planning, recruitment, implementation, and evaluation—be completed within five years. Given the complexities associated with recruitment (which involved contacting

²⁷ Our analysis of improper payments focuses on 2019, the first full calendar year in which beneficiaries were exposed to the benefit offset, as information on improper payments in 2020 was unavailable at the time of the analysis. Additionally, during the COVID-19 pandemic, SSA changed the way that it administered improper payments, particularly not collecting overpayments, which would make interpretation of findings in that year tricky.

over 400,000 SSDI beneficiaries), the implementation period for everyone who enrolled was limited to a minimum of two years.²⁸ Two years might have not been sufficient for people to fully adjust to the new rules. For example, treatment group members became better over time at reporting their earnings in a timely fashion: over the first few months, only about 30 to 40 percent reported earnings timely, whereas by the end of the demonstration 60 to 70 percent reported on time. Timely earnings reports are important as they can limit the potential for improper payments.

B. Size of the incentive

An important question is whether the incentive associated with the offset itself encouraged work. While this offset may be a work disincentive for those earning above the cash cliff amount, it nonetheless might not have acted as an incentive for a broader range of beneficiaries. For people whose benefits were in suspense at the time of enrollment, POD would be an important increase in benefits, but one that would disincentivize work. For such people, the reintroduction of benefits represents a boost to their income – assuming that leisure is a normal good, that would lead to more consumption of leisure (and less work) because of an income effect. Additionally, the offset ramp would represent a work disincentive because of a substitution effect: instead of taking home \$1 for each \$1 of income earned, such beneficiaries would now only take home \$0.50 for each \$1 of income earned with the 2-for-1 offset. That in turn decreases the (relative) price of leisure, thus incentivizing more leisure and less work.

However, several subgroup analyses that try to pinpoint this group for whom the benefit offset represented a theoretical work disincentive do not reveal empirical evidence in line with

²⁸ The first mailings went out in January 2018. The first beneficiaries enrolled in February 2018. To be eligible for the demonstration, one had to submit their enrollment paperwork by December 31, 2018. Given the annual nature of earnings data available, this means that we can only focus on 2019 and 2020 as the two full years of implementation.

this theory (Appendix Tables 1-3). We define subgroups for having benefits suspended in 2018 in the months before enrollment (293 people), having earnings above the SGA amount in 2017 (812 people), and having earnings above the SGA amount at any time since 2014 (1,504 people). These groups capture people for whom there is a positive impact on benefit receipt – all three subgroups have positive impacts that are also differentially significant from the other subgroup that had more limited work activity. Yet there is no statistically significant differential effect on average annual earnings or substantive earnings for any of the three subgroups we consider. For one of the subgroups, the point estimate on average annual earnings is also positive.

Other groups may also not have faced large incentives for work. For example, people with earnings above the TWP amount (\$940 in 2021) but below the SGA amount (\$1,310 in 2021 for non-blind beneficiaries) faced no adjustment to their benefit payments under current rules (since benefits only change when earnings exceed the SGA amount). With the benefit offset, people with earnings in this range would face an additional 50 percent marginal tax rate on earnings given the \$1 for \$2 offset. Thus, such people may face limited incentive to increase work, and could even potentially decrease work activity. Finally, the elimination of the TWP and the grace period, which incentivized work by allowing people to test work activity before their benefits could be reduced, may also decrease work activity.

Within the first two years after enrolling, approximately 30 percent of treatment group members had earnings sufficiently high to have used the benefit offset (Figure 1). About 28 percent of offset users had an offset amount sufficiently low that they experienced reductions in total income relative to current rules.²⁹ To be an offset user, a treatment group member had to

²⁹ The gap between the TWP and SGA amounts was \$350 in 2020, so anyone with an offset amount less than \$175 would have had sufficiently low earnings to fall in this range where POD adversely affected benefits.

earn above the earnings threshold in at least one month after enrollment. Offset users included those who had at least one month of either a partial or a full benefit offset. The median monthly offset amount was \$351. About one-third of offset users (10 percent of all treatment group members) ever had benefits fully offset to \$0.

C. Individual and systemic barriers to employment for people with disabilities

A final factor that might have contributed to the limited impacts observed in our study is the individual and systemic barriers to employment faced by people with disabilities. At the systemic level, people with disabilities face discrimination in the labor market. Correspondence studies like Bellemare et al. (2023) and Ameri et al. (2018) have found that people with disabilities are less likely to get a call back for a job interview. Livermore and Honeycutt (2015) highlight the lower rates of employment for people with disabilities as compared to those without disabilities, while Sundar et al. (2018) document some of the broader barriers that people with disabilities face in employment. Further, the large macroeconomic shock associated with COVID-19 during the second year of the demonstration may have differentially hurt people with disabilities (as it did during the Great Recession; Livermore and Honeycutt 2015).

Our data point to the limited work capacity of individual SSDI beneficiaries specifically. About 90 percent of people in the baseline survey agreed that it was difficult to work because of a physical or mental condition. People also cited other reasons, such as limited skills or a fear of losing benefits or health insurance, as factors that made it difficult to work. Additionally, in the control group, about 55 percent of people did not have any earnings during the two year study period (Table 4). About two-thirds reported not being employed in the past year (Table 4), while ten percent earned above roughly \$15,000 per year (the threshold for substantive earnings in Table 2). Thus, the substantial disabilities that people must have to qualify for SSDI—defined as

an inability to perform any substantial gainful activity—make it difficult for people to work. This limited work capacity is particularly notable given that people who enrolled in the demonstration had a higher predisposition towards work than those who did not enroll. Yet it is important to note that prior research, such as Maestas et al. (2013) and Gelber et al. (2017) suggest a lingering work capacity for many SSDI beneficiaries. Additionally, despite the COVID-19 pandemic, average earnings in the control group were nearly identical in both 2019 (\$4,992) and 2020 (\$4,916).

VII. Discussion

We examined the effect of replacing a cash cliff in SSDI rules, where beneficiaries had benefits reduced to \$0 if their earnings exceeded a key threshold by even \$1, with a benefit offset that reduced benefits by \$1 for every \$2 in earnings. By drastically reducing the marginal tax rate that beneficiaries face on earnings above the cash cliff, this change might have offered greater incentive to work for many. However, for some, the new rules might have reduced the incentive to work.

Overall, we found no impact on earnings, SSDI benefit amounts, or total income. Our estimates are sufficiently precise to rule out substantive changes for the average volunteer: our estimated 95 percent confidence intervals rule out changes of more than 4 percent for benefit amounts and total income, and changes of more than 10 percent for earnings.

A natural question is why this rule change did not impact beneficiary work activity. We present evidence surrounding three potential mechanisms. First, we show that though the demonstration simplified program rules and reduced administrative burden, administrative burden remained wherein people faced challenges with learning costs: limited understanding of the rules, potentially exacerbated by frequent improper payments when beneficiaries work, may

have inhibited work activity. People may also have not had enough time to change their work behavior given the demonstration only ran for two years (the second year also coincided with the start of the COVID-19 pandemic). Second, the incentive may have acted as a work disincentive for people who had previously had benefits suspended because of both income and substitution effects that incentivized more leisure and less work. Finally, broader barriers to employment at both the individual-level and systemic-level may also have prevented people from working. Each of these three elements likely play a role in why replacing the cash cliff with a benefit offset as tested in POD had no effect on employment-related outcomes.

Our findings around administrative burden are important for informing policy, especially given recent substantial interest in the topic such as a recent executive order by President Biden to simplify “customer experiences”.³⁰ Much literature around administrative burden has focused on the ways that administrative burdens, such as through complex application processes, deter and limit participation in programs (e.g., Deshpande and Li 2019; Brodtkin and Majmundar 2010; Finkelstein and Notowidigdo 2019; Moynihan, Herd, and Rigby 2016). Yet the same logic might also dictate that administrative burden from compliance costs can make using the programs more challenging among those qualified (e.g., Homonoff and Somerville 2021; Moynihan, Herd, and Harvey 2015; Heinrich et al. 2021). This in turn might limit whether people optimally take advantage of program rules. Our results do not provide a definitive explanation for the POD findings but fit with the framework developed by Barnes (2021) that highlights redemption costs in understanding challenges associated with using benefits. In particular, the analysis of improper payments, as well as the lack of knowledge of program rules, indicate that administrative burden

³⁰ See Executive Order 14058 from December 13, 2021: <https://www.govinfo.gov/content/pkg/FR-2021-12-16/pdf/2021-27380.pdf>

likely affected POD participants' behavioral responses. Our findings thus indicate a need to reduce administrative burden in programs similar to POD, consistent with the recent executive order.

References

- Anand, Priyanka, Denise Hoffman, John T. Jones, and Siarhei Lukashanets. 2022. “Labor supply response to overpayment notifications: Evidence from Social Security Disability Insurance.” *Contemporary Economic Policy* 40(2): 304–322.
- Brodkin, Evelyn Z., and Malay Majmundar. 2010. “Administrative exclusion: Organizations and the hidden costs of welfare claiming.” *Journal of Public Administration Research and Theory* 20(4): 827–848.
- Deshpande, Manasi, and Yue Li. 2019. “Who Is Screened Out? Application Costs and the Targeting of Disability Programs.” *American Economic Journal: Economic Policy* 11(4): 213–248.
- Finkelstein, Amy, and Matthew J. Notowidigdo. 2019. “Take-up and targeting: Experimental evidence from SNAP.” *The Quarterly Journal of Economics* 134(3): 1505–1556.
- French, Eric, and Jae Song. 2014. “The effect of disability insurance receipt on labor supply.” *American Economic Journal: Economic Policy* 6(2): 291–337.
- Gelber, Alexander, Timothy J. Moore, and Alexander Strand. 2017. “The effect of disability insurance payments on beneficiaries' earnings.” *American Economic Journal: Economic Policy* 9(3): 229–61.
- Gubits, Daniel, David Stapleton, Stephen Bell, Michelle Wood, Denise Hoffman, Sarah Croake, David R. Mann, et al. 2018. “BOND Implementation and Evaluation: Final Evaluation Report.” Final Report submitted to the Social Security Administration. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research.
- Hays, Ron D., Cathy D. Sherbourne, and Rebecca Mazel. 1995. “User's Manual for the Medical Outcomes Study (MOS) Core Measures of Health-Related Quality of Life.” Santa Monica, CA: RAND Corporation. Available at https://www.rand.org/pubs/monograph_reports/MR162.html
- Heinrich, Carolyn J., Sayil Camacho, Sarah Clark Henderson, Mónica Hernández, and Ela Joshi. 2022. “Consequences of administrative burden for social safety nets that support the healthy development of children.” *Journal of Policy Analysis and Management* 12(7): 107–132.
- Herd, Pamela, and Donald P. Moynihan. 2019. *Administrative burden: Policymaking by other means*. Russell Sage Foundation.
- Hock, Heinrich, David Wittenburg, Michael Levere, Noelle Denny-Brown, and Heather Gordon. 2020. “Promoting Opportunity Demonstration: Recruitment and Random Assignment Report.” Final report submitted to the Social Security Administration. Washington, DC: Mathematica.
- Homonoff, Tatiana, and Jason Somerville. 2021. “Program recertification costs: Evidence from SNAP.” *American Economic Journal: Economic Policy* 13(4): 271–98.

- Jolly, Nicholas A., and Kathryn L. Wagner. 2023. "Work-limiting disabilities and earnings volatility." *Labour Economics* 81: 102333.
- Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt." *American Economic Review* 103(5): 1797–1829.
- Moore, Timothy J. 2015. "The employment effects of terminating disability benefits." *Journal of Public Economics* 124: 30–43.
- Moynihan, Donald, Pamela Herd, and Hope Harvey. 2015. "Administrative burden: Learning, psychological, and compliance costs in citizen-state interactions." *Journal of Public Administration Research and Theory* 25(1): 43–69.
- Moynihan, Donald P., Pamela Herd, and Elizabeth Ribgy. 2016. "Policymaking by other means: Do states use administrative barriers to limit access to Medicaid?" *Administration & Society* 48(4): 497–524.
- Nichols, Austin, Jeffrey Hemmeter, and Debra Goetz Engler, eds. 2021. [*Lessons from SSA Demonstrations for Disability Policy and Future Research*](#). Rockville, MD: Abt Press.
- Ruh, Philippe, and Stefan Staubli. 2019. "Financial Incentives and Earnings of Disability Insurance Recipients: Evidence from a Notch Design." *American Economic Journal: Economic Policy*, 11(2): 269–300.
- Schimmel, Jody, David Stapleton, and Jae Song. 2011. "How Common is 'Parking' Among Social Security Disability Insurance Beneficiaries? Evidence from the 1999 Change in the Earnings Level of Substantial Gainful Activity." *Social Security Bulletin*, 71(4): 77–92.
- Weathers, Robert R. II, and Jeffrey Hemmeter. 2011. "The Impact of Changing Financial Work Incentives on the Earnings of Social Security Disability Insurance (SSDI) Beneficiaries." *Journal of Policy Analysis and Management*, 30(4): 708–728.
- Wittenburg, David, Michael Levere, Sarah Croake, Stacy Dale, Noelle Denny-Brown, Denise Hoffman, Rosalind Keith, David R. Mann, Shauna Robinson, Rebecca Coughlin, Monica Farid, Heather Gordon, and Rachel Holzwart. 2022. "Promoting Opportunity Demonstration: Final Evaluation Report." Final report submitted to the Social Security Administration. Washington, DC: Mathematica.

Table 1. Balance across treatment and control groups at baseline

Variable	Treatment group	Control group	<i>p</i> -value
Sex and age			
Female	55.2	54.5	0.473
Mean age (years)	47.4	47.4	0.794
Primary diagnosis			
Neoplasms	2.9	2.9	0.951
Mental disorders	37.9	39.0	0.280
Back or musculoskeletal system	20.1	20.4	0.673
Nervous system disorders	6.5	6.1	0.431
Circulatory system disorders	5.7	6.0	0.553
Genitourinary system disorders	4.3	4.1	0.655
Injuries	3.9	3.8	0.896
Respiratory	1.7	1.7	0.780
Several visual impairments	2.3	2.3	0.848
Digestive system	1.4	1.5	0.632
Other impairments	10.7	9.5	0.049
Beneficiary program characteristics			
Mean SSDI duration (months)	113.3	115.5	0.174
Monthly SSDI benefits (\$)	1035.6	1032.9	0.756
Has representative payee	6.6	7.5	0.106
Concurrent SSI receipt	18.4	17.8	0.442
Employment history			
Recent history of TWP-level earnings	19.0	19.5	0.496
Recent history of SGA-level earnings	14.7	15.3	0.380
Had a Ticket assigned in last four years	12.8	12.0	0.242
Work status at baseline			
Currently employed	24.0	25.1	0.158
Seeking work	23.9	23.5	0.653
Neither employed nor seeking work	52.1	51.4	0.469
Monthly earnings over \$1,000	13.1	13.0	0.976
Expects to work in the next year ^a	61.3	61.0	0.763
Agrees with statement:			
Difficult to work because fear losing disability cash benefits	57.7	57.4	0.790
Difficult to work because of a physical or mental condition	89.5	88.2	0.044
Difficult to work because of unreliable transportation	34.9	33.6	0.196
Difficult to work because caring for children	15.7	16.4	0.415
Difficult to work because don't have needed skills or training	31.8	32.2	0.696
Observations	6,700	3,370	10,070

Source: Authors' calculations using Mathematica's POD recruitment and enrollment system, SSA program records, POD Implementation Data System, and the POD baseline survey.

Note: Unless otherwise noted, all table entries are percentages. The *p*-values in the final column of the table are based on joint tests for differences between the T1, T2, and C groups. These tests compare means for continuous variables, proportions for binary variables, and distributions for multi-valued categorical variables.

Table 2. Impacts on primary outcomes

	(1) Average annual earnings	(2) Substantive earnings	(3) SSDI benefit amount	(4) Average annual income
Treatment	68 (198)	1.0* (0.6)	145 (105)	228 (195)
Control mean	4,954	10.0	11,725	16,548
Observations	10,070	10,070	10,070	10,070

Source: Authors' calculations using SSA program records.

Note: The number for treatment represents an estimate of β from equation (1), representing the estimated impact of being assigned to the treatment group. Standard errors, reported in parentheses, are robust to heteroscedasticity. All monetary values are in 2019 dollars.

Substantive earnings is an indicator variable for whether the beneficiaries' average earnings over 2019 and 2020 exceeds the annualized SGA amount over the full two-year period. All outcomes are measured over the 2019 and 2020 calendar years. The exception is that SSDI benefit amounts are measured over the two years after POD enrollment.

***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Table 3. Impacts on primary outcome, by employment status at time of enrollment

	(1) Average annual earnings	(2) Substantive earnings	(3) SSDI benefit amount	(4) Average annual income
Panel A. Employed				
Treatment	-462 (647)	0.9 (1.8)	330 (266)	-101 (596)
Control mean	13,581	29.0	10,337	23,608
Observations	2,341	2,341	2,341	2,341
Panel B. Not employed				
Treatment	298* (156)	1.2** (0.5)	75 (110)	383** (171)
Control mean	2,224	4.0	12,164	14,314
Observations	7,729	7,729	7,729	7,729
Panel C. Differential impact <i>p</i>-value				
Treatment	0.253	0.892	0.375	0.436

Note: The number for treatment represents an estimate of β from equation (1), representing the estimated impact of being assigned to the treatment group, for those within each subgroup. Standard errors, reported in parentheses, are robust to heteroscedasticity. The differential impact *p*-values in Panel C come from a test of whether the impact estimate for those who were employed at POD enrollment is equal to the impact estimate for those who were not employed at POD enrollment. All monetary values are in 2019 dollars. Substantive earnings is an indicator variable for whether the beneficiaries' average earnings over 2019 and 2020 exceeds the annualized SGA amount over the full two-year period. All outcomes are measured over the 2019 and 2020 calendar years. The exception is that SSDI benefit amounts are measured over the two years after POD enrollment. Those with missing employment status (95 people) at the time of enrollment are assumed to be not employed as that was the more common response.

***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Table 4. Impacts on supplemental outcomes

	Control group mean	Estimated impact	Standard error	Sample size
Employment-related outcomes				
Any employment in past year	33.5	1.0	1.0	7,842
Employed or actively searching for a job	50.9	2.6**	1.1	7,867
Any positive earnings (SSA program records)	45.0	1.0	0.9	10,070
Hours worked per week at most recent job	8.0	0.4	0.3	7,842
Any fringe benefits offered at most recent job	17.4	1.5*	0.8	7,842
Applied for VR services in first two years after enrollment (VR program records)	2.8	1.3***	0.4	10,070
Received VR services in first two years after enrollment (VR program records)	4.0	0.7*	0.4	10,070
Health and health insurance related outcomes				
Physical health aggregate score	33.9	0.0	0.3	6,971
Mental health aggregate score	39.3	-0.0	0.3	6,971
Has any health insurance coverage	98.5	0.2	0.3	7,732

Source: Authors' calculations using SSA program records.

Note: The estimated impact represents an estimate of β from equation (1). Unless otherwise noted, all table entries are percentages for means or percentage points for impact estimates. Unless otherwise noted, all data are from the year two follow-up survey. Data from the follow-up survey can be missing owing to item-level non-response. Data from the follow-up survey are weighted using survey non-response weights to account for the people who were sent the survey but did not complete it. The administrative data includes all initial participants in the demonstration. Standard errors, reported in parentheses, are robust to heteroscedasticity. ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

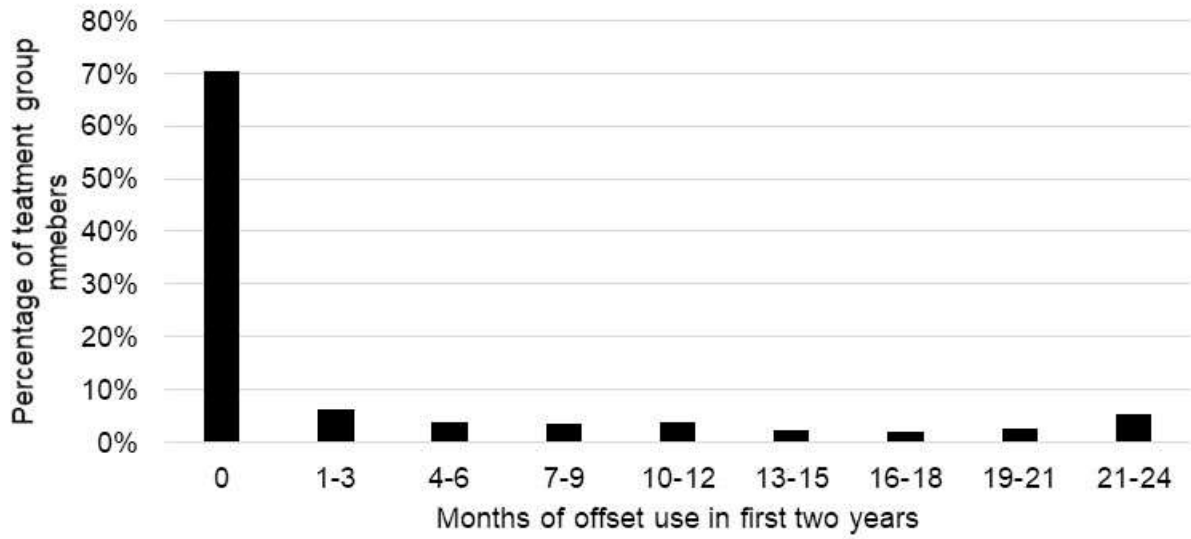
Table 5. Understanding of SSDI work rules

	One year after enrollment	Two years after enrollment
Panel A. Treatment group members		
Understand that there is no Trial Work Period	34.0	34.9
Understand the termination rules	34.7	33.9
Understand that benefits can be reduced if monthly earnings are sufficiently high	49.0	46.1
Number of responses	2,635	5,054
Panel B. Control group members		
Understand that there is a Trial Work Period	28.0	28.0
Understand that benefits can be terminated if earnings are too high	44.0	43.6
Number of responses	1,438	2,803

Source: POD two-year follow-up survey

Note: The following three questions assessed the understanding of treatment group members about POD rules: (1) Under POD, do you have a TWP where your benefits remain unchanged regardless of your earnings? (2) Under the POD rules, do your benefits ever terminate if your earnings are too high? (3) Under POD, are your benefits reduced at any time if your monthly earnings are above a level that SSA set for POD? The correct answer to the second question differed depending on the version of the POD rules. The following two questions assessed the understanding of current SSDI rules by control group members: (1) Under current SSDI rules, do you have a Trial Work Period where your benefits remain unchanged regardless of your earnings? (2) Under current SSDI rules, do your benefits ever terminate if your earnings are too high? Beneficiaries had the option to answer that they did not know the answer.

Figure 1. Distribution of months of offset use in first two years



Source: Authors' calculations using POD Implementation Data System.

Note: Includes all 6,700 treatment group members.

Appendix Table 1. Impacts on primary outcome, by suspense status at time of enrollment

	(1) Average annual earnings	(2) Substantive earnings	(3) SSDI benefit amount	(4) Average annual income
Panel A. Benefits suspended				
Treatment	502 (2,563)	-0.5 (5.4)	2,929*** (747)	3,475 (2,278)
Control mean	26,550	65.0	4,094	30,726
Observations	293	293	293	293
Panel B. Benefits not suspended				
Treatment	272 (182)	1.6*** (0.6)	-49 (103)	246 (186)
Control mean	4,039	7.7	12,063	15,947
Observations	9,777	9,777	9,777	9,777
Panel C. Differential impact <i>p</i>-value				
Treatment	0.929	0.892	0.000	0.158

Note: The number for treatment represents an estimate of β from equation (1), representing the estimated impact of being assigned to the treatment group, for those within each subgroup. Standard errors, reported in parentheses, are robust to heteroscedasticity. The differential impact *p*-values in Panel C come from a test of whether the impact estimate for those whose benefits were suspended at POD enrollment is equal to the impact estimate for those whose benefits were not suspended at POD enrollment. All monetary values are in 2019 dollars. Substantive earnings is an indicator variable for whether the beneficiaries' average earnings over 2019 and 2020 exceeds the annualized SGA amount over the full two-year period. All outcomes are measured over the 2019 and 2020 calendar years. The exception is that SSDI benefit amounts are measured over the two years after POD enrollment. Benefit suspense is defined as having benefits suspended for work in any month in 2018 in the months prior to being enrolled.

***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Appendix Table 2. Impacts on primary outcome, by 2017 earnings status at time of enrollment

	(1) Average annual earnings	(2) Substantive earnings	(3) SSDI benefit amount	(4) Average annual income
Panel A. Any months in 2017 with SGA-level earnings				
Treatment	-703 (1,286)	-1.2 (3.3)	1,178** (483)	465 (1,160)
Control mean	17,475	40.6	8,022	25,322
Observations	812	812	812	812
Panel B. No months in 2017 with SGA-level earnings				
Treatment	226 (180)	1.4*** (0.6)	8 (105)	251 (185)
Control mean	3,686	6.9	12,115	15,659
Observations	9,258	9,258	9,258	9,258
Panel C. Differential impact p-value				
Treatment	0.474	0.431	0.018	0.856

Note: The number for treatment represents an estimate of β from equation (1), representing the estimated impact of being assigned to the treatment group, for those within each subgroup. Standard errors, reported in parentheses, are robust to heteroscedasticity. The differential impact p -values in Panel C come from a test of whether the impact estimate for those had any months in 2017 with SGA-level earnings is equal to the impact estimate for those who had no months in 2017 with SGA-level earnings. All monetary values are in 2019 dollars. Substantive earnings is an indicator variable for whether the beneficiaries' average earnings over 2019 and 2020 exceeds the annualized SGA amount over the full two-year period. All outcomes are measured over the 2019 and 2020 calendar years. The exception is that SSDI benefit amounts are measured over the two years after POD enrollment.

***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Appendix Table 3. Impacts on primary outcome, by recent earnings status at time of enrollment

	(1) Average annual earnings	(2) Substantive earnings	(3) SSDI benefit amount	(4) Average annual income
Panel A. Any months since 2014 with SGA-level earnings				
Treatment	-657 (865)	0.7 (2.3)	1,168*** (341)	567 (771)
Control mean	13,985	30.6	9,525	23,239
Observations	1,504	1,504	1,504	1,504
Panel B. No months since 2014 with SGA-level earnings				
Treatment	204 (175)	1.1** (0.5)	-38 (108)	173 (184)
Control mean	3,321	6.3	12,139	15,338
Observations	8,566	8,566	8,566	8,566
Panel C. Differential impact p-value				
Treatment	0.329	0.871	0.001	0.619

Note: The number for treatment represents an estimate of β from equation (1), representing the estimated impact of being assigned to the treatment group, for those within each subgroup. Standard errors, reported in parentheses, are robust to heteroscedasticity. The differential impact p -values in Panel C come from a test of whether the impact estimate for those had any months since 2014 with SGA-level earnings is equal to the impact estimate for those who had no months since 2014 with SGA-level earnings. All monetary values are in 2019 dollars. Substantive earnings is an indicator variable for whether the beneficiaries' average earnings over 2019 and 2020 exceeds the annualized SGA amount over the full two-year period. All outcomes are measured over the 2019 and 2020 calendar years. The exception is that SSDI benefit amounts are measured over the two years after POD enrollment.

***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.