Lessons from SSA Demonstrations for Disability Policy and Future Research

Edited by
Austin Nichols  ■  Jeffrey Hemmeter  ■  Debra Goetz Engler
Overview

Over the past several decades, the Social Security Administration has tested many new policies and programs to improve work outcomes for Social Security Disability Insurance beneficiaries and Supplemental Security Income recipients. These demonstrations have covered most aspects of the programs and their populations. The demonstrations examined family supports, informational notices, changes to benefit rules, and a variety of employment services and program waivers.

A “State of the Science Meeting,” sponsored by the Social Security Administration and held on June 15, 2021, commissioned papers and discussion by experts to review the findings and implications of those demonstrations.

A subsequent volume—Lessons from SSA Demonstrations for Disability Policy and Future Research—collects the papers and discussion from that meeting to synthesize lessons about which policies, programs, and other operational decisions could provide effective supports for disability beneficiaries and recipients who want to work. This PDF is a selection from that published volume. References from the full volume are provided.

Suggested Citations


Chapter 4

The Return to Work in Disability Programs: What Has Been Learned and Next Steps

Jesse Gregory  
*University of Wisconsin*  
Robert A. Moffitt  
*Johns Hopkins University*

The United States has two major government programs for individuals with disabilities. The first, the Social Security Disability Insurance (SSDI) program, provides cash benefits and health insurance in the form of Medicare to individuals younger than age 66 or 67 who meet a test for severe disability and who have sufficient past earnings in jobs covered by the Social Security system.\(^1\) The amount of the monthly cash benefit is determined by the level of past earnings. The second, the Supplemental Security Income (SSI) program, is means-tested and provides cash payments and Medicaid coverage to individuals who meet the same disability test as in SSDI, but who have low income and assets and hence qualify under a current means test rather than under a work-history test.\(^2\) The basic monthly benefit level depends on marital status and living arrangements, unearned income, and other factors, but not on past work history.\(^3\)

Both SSI and SSDI are intended to provide support only to those who have no or little capacity to work, and the history of the discussions even at the time of creation of these programs reveals a concern with work disincentives (Berkowitz 2013, 2020). An emphasis on rehabilitation has always been a central focus of the programs. However, as the caseloads of both programs began to rise in the 1980s and accelerated in the 1990s and 2000s (but have recently peaked and begun to fall), discussions of how to promote returns to work among SSDI beneficiaries and SSI recipients increased. This concern led to considerable discussion and congressional attention and academic research but also led the Social Security Administration (SSA) to conduct a number of demonstrations, almost always using randomized control trial methods, to test possible reforms to the programs to increase work, employment, and earnings.

---

1. SSDI is a social insurance program, where the premiums are the tax contributions made from covered earnings. We are referring to worker beneficiaries, with which this chapter is concerned, not disabled adult children or disabled widows or widowers.

2. We will only be concerned with the SSI program for adults, not children or the elderly. Elderly and child SSI recipients have different eligibility rules.

3. It is also possible to receive benefits from both programs (these are called “concurrent beneficiaries”). The programs both have work incentives and access to Vocational Rehabilitation and the Ticket to Work programs that we discuss below.
Though both SSDI and SSI intend to provide benefits only to those with severe disabilities and not partial disabilities, some concurrent beneficiaries have some ability to work—residual functional work capacity (often called residual work capacity or “work capacity” for short)—although the number with the capacity for substantial work levels is uncertain. There is widespread agreement that those with capacity to work should be encouraged to use that capacity through labor market activity, and that those with substantial capacity who can leave the program should be encouraged to do so.4

One major concern about the structure of the current SSDI program is that the rules of the program governing work might provide meaningful work disincentives to those with such residual capacity. The rules follow from the philosophy of the program, which is to divide the population into those who can work and those who cannot and to provide benefits only to the latter—to provide a safety net for those unable to work at a minimally self-sustaining level (Substantial Gainful Activity, or SGA). Several steps in the determination process are designed to make that division of the applicant pool into the two groups, but there are inevitably errors in that determination. An individual’s work capacity can change over time, as well. Once in the program, an individual is designated to be able to work if they demonstrate that they do, in fact, have residual work capacity through actual work and earnings or if medical determinations show that they no longer meet the disability standard for the program. Therefore, the program rules are designed to test whether an individual’s disability prevents them from performing work considered SGA.5

The SSI program provides less direct financial work disincentive than does SSDI, imposing a 50 percent benefit reduction rate (BRR) on earnings (after a disregard and deductions) rather using earnings per se as a test of disability (although both provide work programs of other types and special incentives for the blind). However, few recipients of SSI work at high levels, similar to the SSDI program, and the desire to increase employment and labor force engagement among those with residual work capacity is similar to that for SSDI beneficiaries.

In the first section of this chapter, we review the rules of the SSDI and SSI programs with a focus on those governing work and earnings. The second section discusses conceptual issues that should influence thinking about how those rules might affect individual motivations to work. The third section of this chapter reviews the evidence from a number of SSA demonstrations. The demonstrations reviewed are only those designed for individuals who are already beneficiaries or recipients of one of the two programs or have applied, and not demonstrations that attempt early interventions prior to application (those are covered by Hollenbeck [2021]). The fourth

4 Both programs explicitly state that encouraging beneficiaries and recipients to work if they can is one of the goals of the programs (SSA 2020b, 6; SSA 2020h, 1).
5 There have been many proposals for expanding the SSDI program, or creating new ones, to cover the partially disabled. See Maestas (2019) for one such recent proposal. We consider such proposals briefly in the final section of this chapter.
section draws some general lessons learned from the demonstrations. The fifth section suggests directions for new demonstrations and demonstration designs that might be considered. A short summary concludes the chapter.

**STRUCTURE OF THE SSDI AND SSI PROGRAMS**

**Social Security Disability Insurance Program (SSDI)**

The SSDI program, begun in 1956, had slightly more than 8 million disabled workers in 2019 (SSA 2020b). About 30 percent establish eligibility on the basis of an intellectual disability or other mental disorder, 34 percent on the basis of a musculoskeletal system or connective tissue disability, and the rest on the basis of a range of other impairments. The caseload has grown for several decades, with disproportionate growth among those with mental and musculoskeletal impairments. To the extent that residual work capacity differs by impairment type, these trends could affect aggregate caseload residual capacity. The caseload peaked in 2014 and has declined since, partly from demographic trends and partly from administrative changes (Maestas 2019).

There is a large literature on the sources of the rapid caseload growth, most of which is outside the scope of our review. However, one suggested contributor to growth is the long-term decline in the demand for unskilled labor in general—a trend at least since the 1980s—with a consequent decline in average wages paid in the private sector for low-skilled jobs. Autor and Duggan (2000) argue that the decline in those wages has been one of the forces pushing up caseloads, as individuals with disabilities tend to work at such jobs. If so, the implication is that increasing employment of SSDI beneficiaries or increasing the rate of exit from SSDI to employment could have become more difficult.

Exits from SSDI because of a successful return to work are rare. Though almost 10 percent of beneficiaries have had their benefits terminated in recent years, most of these are the result of death or reaching the full retirement age. Only 0.7 percent of SSDI worker beneficiaries terminated benefits from a successful return to work in 2019 (SSA 2020b, Table 50). Those with mental impairments had higher rates of termination whereas those with musculoskeletal impairments had lower rates. Some beneficiaries, as described below, have benefits withheld for a month because of high earnings but are not terminated. Another 0.7 percent of beneficiaries experienced such

---

6 This constitutes 89 percent of the caseload, the rest consisting of disabled adult children and disabled widows and widowers.

7 The growth in the share of SSDI caseload with mental and musculoskeletal impairments corresponds with an increase in the prevalence of these impairments among the broader US population with disabilities (Mojtabai 2011; Freburger et al. 2009).

8 Longitudinal studies show higher rates of exit (Liu and Stapleton 2011).
withholding in December 2019. Those with mental impairments had about the same rate whereas those with musculoskeletal disorders had a lower rate.

Application for SSDI involves a five-step process that begins with a determination of whether the applicant is engaged in SGA, which in 2021 was $1,310 per month for a nonblind applicant and $2,190 for a blind applicant. After also determining whether the disability is severe, the individual’s disability must be found in a list of impairments. Then an assessment is made of whether the individual can return to their pre-disability job or to any other job in the economy. Denials can be appealed, and the ultimate decision in individual cases can sometimes take several years. There have been many proposals to reform elements of this process, but most are not directly related to work disincentives other than the SGA criterion itself (we mention an exception in the next section).

After an SSDI award, a number of work-related rules are present in the program. The first, called the Trial Work Period (TWP), begins if earnings are above a threshold, allowing the beneficiary to experience no reduction in benefits for up to nine months within a 60-month window. Those who complete a TWP have a 3-month Grace Period the first month the beneficiary works SGA, followed by an EPE, which is a 36-month reentitlement period during which benefits are withheld in months with earnings at or above the SGA level but fully paid in months with earnings below the SGA level. After the 36-month period elapses, any subsequent month with earnings above the SGA level results in benefit termination and exit from the program (but with extended Medicare coverage). An expedited reinstatement application is allowed if the individual becomes unable to work again within the next 60 months. In the 2014–2018 period, about 2 to 3 percent of beneficiaries per year used the TWP and about half of those completed a TWP (SSA 2020a, Tables 1 and 58).

There are a large number of other work-related SSDI rules that have not been directly studied (i.e., have not been the subject of randomized trials). These include rules governing unsuccessful work attempts, deductions for impairment-related work expenses, continued payments under certain circumstances for those engaged in Vocational Rehabilitation (VR), provisions for extended Medicare coverage after completion of the TWP, options to buy Medicare for some period after benefit termination, and a variety of services intended to help those who wish to return to work.

---

9 There are a number of other calculations of earnings amounts prior to application of the TWP rule. See the Red Book (SSA 2020e) at https://www.ssa.gov/redbook/.

10 Again, the rates could be higher in longitudinal statistics.

11 These other rules are discussed in SSA’s Red Book (https://www.ssa.gov/redbook/eng/resources-supports.htm).
Supplemental Security Income Program (SSI)

Enacted by Congress in 1972, the SSI program replaced a state-level system of support for the low-income, aged, blind, and disabled population with a federal program with nationwide benefits and eligibility criteria. The definition of disability used by the SSI program for initial award consideration is the same as that for the SSDI program, but SSI also requires that income and assets fall below certain specific levels, as well as having a number of citizenship requirements.\(^\text{12}\)

After an award, SGA does not come into play because under SSI there is no direct post-award earnings test for being disabled as there is for SSDI.\(^\text{13}\) SSI instead operates much more like a traditional transfer program, as it is intended to do, with a maximum payment made to those with zero earnings and with a reduced benefit as earnings rise (after certain deductions). The BRR is 50 percent, not far from the rate in many other means-tested transfer programs. There are also a number of other services designed to encourage a return to work, some shared with SSDI and some only for SSI.\(^\text{14}\) Unearned income (with certain exceptions) reduces benefits dollar-for-dollar, following from the philosophy of the program as means-tested rather than as social insurance. This treatment allows SSI recipients to receive SSDI (termed concurrent beneficiaries) as long their SSDI benefit is not so large as to make them ineligible for an SSI payment after disregards. Finally, a number of states also have supplemental SSI programs.\(^\text{15}\)

The percentage of blind and disabled SSI recipients ages 18 to 64 who work was 7.9 percent in 2007 and fell to 7.0 percent in 2019;\(^\text{16}\) 2019 employment rates are higher for those with autistic disorders (17 percent) and intellectual disorders (12 percent), but are close to zero for many other diagnostic groups. Employment rates also are highest among recipients in their 20s and 30s. Exit rates because of work are low: of those receiving benefits in 2018, only 1.4 percent had achieved high enough earnings in 2019 to no longer be eligible for the program.\(^\text{17}\)

\(^{12}\) SSI and SSDI have different look-back periods for back benefits after an award is made. The earnings limits increase with inflation, but the asset limits do not and have been held constant in nominal terms for many years. Those admitted to SSI are generally eligible for Medicaid on the basis of their income.

\(^{13}\) What are termed “Section 1619 waivers” allow recipients to work above the SGA level.


\(^{15}\) See Duggan, Kearney, and Rennane (2016) for a description of the range of state programs.

\(^{16}\) See SSA (2020h, Table 42) and earlier years back to 2007. Figures earlier than 2007, but for work rates for all SSI recipients regardless of age, peaked in 2000 and have been falling since then (SSA 2020h, Table 40).

\(^{17}\) This is for recipients in 1619(b) status. See SSA (2020h, Table 44: 102,867 divided by 6,971,113). This earnings level is generally above the SGA level. For example, in 2020 an individual living alone with no deductions other than earnings and with no income other than earnings could earn up to $1,436 in a month before losing eligibility (the SGA threshold is $1,220). Longitudinal statistics could show different exit rates.
CONCEPTUAL ISSUES

In this section, we discuss conceptual issues surrounding providing work support and work incentives to recipients of the SSI program and beneficiaries of the SSDI program and to the severely disabled population in general. We draw upon several decades of thinking and research on this topic from the economic literature on work incentives of transfer programs for the needy in general, and where those apply to recipients of disability benefits, as well as from the research specifically on SSDI and SSI. There are many general lessons from this literature that should help guide thinking for these programs, but also a number of dimensions that differ for the SSDI and SSI programs.\footnote{See Moffitt (2016) for reviews of the major US means-tested transfers and their effects on work effort.}

The conventional view of employment decisions, and of decisions about working part-time, full-time, or some other hours of work, is that individuals compare the value of working at different levels to the value of not working. The value of working includes financial compensation, fringe benefits, and health insurance provision, for example. And most individuals also place non-economic value on working because it improves an individual’s feeling of self-worth and contributing to society. The value of not working includes the value of income available in that status, as well as the value to the individual of the reduction of time spent at work (or more generally, the disutility of work). Health considerations always play a role in the disutility of work, and that is an obvious element integral to labor force decisions of those with disabilities. Transportation costs and other fixed costs of working also play a role in the value of working, as does the availability of accommodation. The availability of income and care support from others in the household is important to the value of not working.

Lack of knowledge of the labor market for individuals with disabilities could also play a role insofar as individuals who do not work may not know the types of jobs that they might have open to them, and it may take substantial time and effort to search for jobs to acquire that information. On the labor demand side, the number of job openings that are suitable to the individual is important, as is the state of the business cycle.\footnote{Discrimination against individuals with disabilities may affect the jobs available, and racial discrimination may also hinder job opportunities for those with disabilities.} There are also dynamic considerations, the two most important being the positive impact of working on future earning power (human capital)—most important for younger workers—and the negative impact of approaching retirement on the incentive to invest in acquiring new skills—a consideration important for older workers.\footnote{There may be other dynamic considerations.}

All these considerations can be safely assumed to be part of the benefits calculus of individuals with disabilities. Medical conditions increase the disutility of working, and fixed costs of work as well as transportation costs are likely to be important, as is
lack of information about jobs that are suitable and provide some degree of accommodation. These factors are likely to be more important for individuals with disabilities than for individuals out them, and all will tend to discourage work. As for SSDI and SSI, if an individual is not working, those programs provide cash support and health insurance, both of which also generate work disincentives by providing support to those who do not work. A major problem with assessing the effects of SSDI and SSI rules on labor force decisions of their beneficiaries and recipients is that the general responsiveness to the factors affecting those decisions is not known from the research literature. The basic building blocks of the economic model outlined above are the concepts of substitution and income effects. The “substitution effect” refers to how much more individuals will work if the financial gain from working increases, and the “income effect” refers to how much less individuals will work if they are given higher benefits for not working. When the “dynamic” considerations mentioned above are considered, the concept of discount rates (how individuals weigh future outcomes), and the returns to human capital from the types of jobs that individuals with disabilities have open to them, are relevant to individual decisions. Although substantial existing evidence estimates these factors for those without disabilities, estimates are largely absent from the research literature for those with disabilities. However, many of the demonstrations we will review in the next section have suggestive evidence on some of these factors.

Returning to the general research literature on work incentives in transfer programs, most economists view the problem as one that requires finding a balance between providing support for a needy population in the form of cash, medical care, or subsidies of various kinds, and the preservation of at least some incentives to work. The goal of the programs is to continue to provide support to those who need it, while encouraging and assisting those with residual work capacity to use that capacity. The role of research in this enterprise is to try to identify mechanisms that reduce the magnitude of this central tradeoff by finding reforms that make it easier to achieve that balance.

This tradeoff is handled differently in SSDI and SSI. In SSDI, beneficiaries ultimately deemed capable of working at the SGA level are regarded as not disabled and experience benefit termination. But in SSI, working above the SGA level is allowed; recipients with higher levels of earnings are considered eligible, as long as they are still judged to meet the medical definition of disability and meet the countable income test. In SSDI, beneficiaries who consistently work above the SGA level are only eligible for a limited time.

When thinking about ways to manage the tradeoff, a well understood issue in the existing literature is that transfer programs with high BRRs provide work disincentives to recipients, but lowering those rates does not by itself lead to program exit. Indeed,

---

21 Autor and Duggan (2007) argue that income effects for those with disabilities can be estimated from the Department of Veterans Affairs program.
the original Negative Income Tax proposals by Milton Friedman were intended to allow recipients to keep receiving benefits after beginning to work, and that this supplementation would be permanent, not temporary. The SSI program accepts that tradeoff to a greater degree than the SSDI program does.²² The evidence we will review in the next section on SSDI suggests that allowing beneficiaries to work above the SGA level, but only temporarily (because, eventually, they will be terminated from the program), might not have a very large impact on labor force engagement.

Another principle from the literature on low BRRs is that reducing those rates has ambiguous effects on labor supply because, by extending upward the range of earnings over which benefits can be received, the BRR is increased for some individuals (Hoynes and Moffitt 1999; Moffitt 1992b).²³ For individuals with earnings that fall beyond the initial phase-out region but fall within the expanded phase-out region, a substitution effect will tend to reduce labor supply. Additionally, an income effect will tend to reduce labor supply for all individuals whose benefits increase at their initial earnings level. This has a direct application to benefit cliffs such as that in the SSDI program for the SGA level. Everyone dislikes benefit cliffs, but smoothing them out has ambiguous effects on average levels of labor supply: Increases in benefits tend to reduce labor supply through income effects, and some individuals face lower BRRs whereas others face higher rates, all with an ambiguous sign on the net effect. Smoothing out benefit cliffs implies increasing some beneficiaries’ work levels and reducing them for others, which implies a tradeoff across individuals—we must consider whether a higher-earnings group that gets the smoothing is more worthy of support than the group not subject to the smoothing.²⁴

These simple lessons from other programs take a different form in SSDI, with its work rules in the form of the TWP and the EPE. The TWP imposes a zero BRR, the EPE reentitlement period imposes a benefit cliff for earnings at the SGA level, and benefits are terminated for earnings above the SGA level in the post-reentitlement EPE period (ignoring the Grace Period). Altering the TWP by imposing a nonzero BRR above the TWP threshold or smoothing the benefit cliff during the reentitlement period would both have ambiguous net effects on work levels. But, unlike means-tested programs where high enough levels of work result in fairly immediate termination from the program, in SSDI the beneficiary has up to eight years before termination occurs (60 months for the TWP but which can occur in any window after receipt begins, and 36 months for the EPE reentitlement period). A beneficiary therefore has quite a long time to establish skills and work patterns before having their benefits terminated. Probably also important is the uncertainty that individuals have of their own earning ability and how difficult it will be to sustain working if their medical

²² Administrative costs in conducting large numbers of work continuing disability reviews (CDRs) in SSDI for very small supplements could play a role here.

²³ Levy (1979) was the first paper to note this result in the welfare program literature.

²⁴ However, lowering BRRs often unambiguously increases the probability of being employed at all.
conditions remain as barriers, although the purpose of the TWP and EPE is to allow the beneficiary to test out their ability to work. The usual presumption is that, in the face of uncertainty, most individuals are risk averse and are less likely to undertake actions such as working in the face of that uncertainty. These considerations are absent in SSI, which does not have an earnings test per se to determine whether disability continues.

A consideration sometimes mentioned in the literature on other transfer programs is that work can increase human capital and therefore earning power in the labor market, which could mean that short-term incentives to work could have a longer-run impact by increasing exit rates from the program. However, most unskilled jobs have relatively low rates of return to human capital and so any effect of this kind is likely to be small.²⁵

Another “dynamic” consideration results from the unique feature of SSDI and SSI (not present in other transfer programs) represented by their extensive application process, with its determination of medical condition and work capacity. That the determination period can be quite long for some applicants initially denied (although not the majority), and that individuals often do not risk working even below the SGA level during that period, could reduce long-term employability and even the desire for work (Autor et al. 2015). It is therefore possible that the features of the application process have effects on labor force engagement among SSDI beneficiaries and SSI recipients quite apart from the effects of BRRs, VR, and the other central work supports provided to recipients after an award.

REVIEW OF SSA DEMONSTRATIONS

The changes in policies, programs, and services to promote SSDI beneficiaries’ and SSI recipients’ return to work that SSA demonstrations have studied fall in four broad categories: (1) those that operate through changes to financial incentives, (2) those that involve modifications in the use of VR services or other direct employment services, (3) those that offer new employment programs specifically aimed to individuals with mental impairments, and (4) those that modify health insurance coverage. We divide our survey of SSA demonstrations based on this categorization, but recognizing that some demonstrations involve multiple interventions to some degree.

²⁵ And, in fact, encouraging employment can reduce future human capital to the extent it reduces time that would otherwise be spent in education or formal training, as noted by Heckman, Lochner, and Cossa (2003). The discouraging effect on education is actually one reason why the Ticket to Work program doesn’t include all those of working age (i.e., age 16 and older). However, education and training could be less important for the older SSDI population.
### Exhibit 4.1. Demonstrations Reviewed

<table>
<thead>
<tr>
<th>Name</th>
<th>Status</th>
<th>Years</th>
<th>Population</th>
<th>Intervention</th>
<th>Results to Date</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Financial Incentives</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benefit Offset National Demonstration (BOND)</td>
<td>Completed</td>
<td>2011–2015a</td>
<td>• SSDI beneficiaries • Concurrent beneficiaries</td>
<td>$1 for $2 benefit offset for annual earnings</td>
<td>Increased benefits, differently signed effects on earnings: above BYA (+), above 2 × BYA (−)</td>
</tr>
<tr>
<td>Promoting Opportunity Demonstration (POD)</td>
<td>In progress</td>
<td>2018–</td>
<td>• SSDI beneficiaries</td>
<td>Elimination of TWP/EPE, $1 for $2 benefit offset</td>
<td>Results not yet available</td>
</tr>
<tr>
<td>State Partnership Initiative (SPI)</td>
<td>Completed</td>
<td>1999–2004</td>
<td>• SSDI beneficiaries • SSI recipients • Others with disabilities</td>
<td>Combinations of case management, job training, employer outreach, and other employment services</td>
<td>No experimental results to report due to lack of credible comparison groups</td>
</tr>
<tr>
<td>State Partnership Initiative (SPI) SSI Work Incentives Demonstration Project</td>
<td>Completed</td>
<td>2001–2004</td>
<td>• SSI recipients • Concurrent beneficiaries</td>
<td>$3 for $4 earnings exemption, allowed asset accumulation, suspended certain CDRs</td>
<td>No experimental results to report due to lack of credible comparison groups</td>
</tr>
<tr>
<td><strong>Vocational Rehabilitation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project NetWork</td>
<td>Completed</td>
<td>1992–1995</td>
<td>• SSDI beneficiaries • SSI applicants and recipients</td>
<td>Case management services</td>
<td>Increased service receipt, small short-run increases in earnings</td>
</tr>
<tr>
<td>Ticket to Work (TTW)b</td>
<td>Completed</td>
<td>2002–</td>
<td>• SSDI beneficiaries • SSI recipients • Concurrent beneficiaries</td>
<td>Employment services, financial incentive to service providers</td>
<td>Increased service receipt, entry of services providers</td>
</tr>
<tr>
<td>Ohio Direct Referral Demonstration (ODRD)</td>
<td>In progress</td>
<td>2020–</td>
<td>• Young SSDI and SSI applicants and beneficiaries/recipients</td>
<td>Direct referral to VR</td>
<td>Results not yet available</td>
</tr>
<tr>
<td>Name</td>
<td>Status</td>
<td>Years</td>
<td>Population</td>
<td>Intervention</td>
<td>Results to Date</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>----------</td>
<td>--------------------</td>
<td>-------------------------------------------------</td>
<td>--------------------------------------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>Transitional Mental Impairments</td>
<td>Completed</td>
<td>1985–1987</td>
<td>SSI recipients with intellectual disability</td>
<td>Job placements, specialized OJT, postplacement support</td>
<td>Increased employment and earnings</td>
</tr>
<tr>
<td>Transitional Employment for Mental Impairments Training Demonstration (TETD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental Health Treatment Study (MHTS)</td>
<td>Completed</td>
<td>2006–2010</td>
<td>SSDI and concurrent beneficiaries with schizophrenia or affective disorder</td>
<td>Comprehensive set of services including IPS</td>
<td>Increased employment and earnings</td>
</tr>
</tbody>
</table>
| Supported Employment Demonstration (SED)   | In progress | 2017– | SSDI and SSI denied applicants with mental impairments | 1. Basic services including IPS  
2. Plus services of a nurse coordinator | Results not yet available |
2. Plus other services | Increased employment in 2nd year, none by 3rd year |

Key: BYA=BOND Yearly Amount. TWP=Trial Work Period. EPE=Extended Period of Eligibility. CDR=continuing disability review. IPS=Individual Placement and Support. OJT=on-the-job training. VR=Vocational Rehabilitation.

Some beneficiaries will continue to receive benefits into 2022.

Ticket to Work was not an SSA demonstration. We include it in this review because it is an important component of recent policy facilitating return to work and because its intentionally staggered rollout provides quasi-experimental policy variation that was the focus of evaluations of the program.

**Demonstrations Focused on Financial Incentives**

**Benefit Offset National Demonstration (BOND)**

The largest demonstration to date studying the impact of removing financial work disincentives on return-to-work outcomes was BOND, which was mandated by Congress in the Ticket to Work and Work Incentives Improvement Act of 1999 (Ticket Act).²⁶ Beginning in 2011, BOND used random assignment to select treatment and control groups from a large nationally representative cross section of the SSDI beneficiaries under age 60. All SSDI beneficiaries (including those concurrent with SSI) between ages 20 and 59 from sampled Area Offices and not in another SSA demonstration were included in the main BOND sample, which was called Stage 1 (N=77,101 treatment group, N=891,429 control group). Different from many

---

²⁶ Because of its importance, we spend more time reviewing it than the others that follow.
demonstrations, this meant that the BOND Stage 1 sample was not limited to volunteers. A smaller sample of SSDI-only beneficiaries predicted to be most likely to use the offset were recruited to volunteer for a second intervention, called Stage 2 (treatment group \(N=7,895\), control group \(N=4,849\)). All participants in Stage 2 were volunteers. Treatment or control status was randomly assigned to participants in both stages.\(^{27}\)

The treatment groups in both Stage 1 and Stage 2 received a modification to the normal SSDI benefit formula that removed the cash cliff faced by SSDI beneficiaries who have completed a TWP. Under the normal SSDI program rules, following the TWP, monthly earnings above the SGA level result in the loss of SSDI benefits. BOND did not alter the rules governing the TWP itself. Instead, under the modified rules for the BOND treatment groups, following the TWP, earnings above a related earnings threshold for annual earnings called the BOND Yearly Amount (BYA) resulted in a $1 reduction in benefits for each $2 of additional earnings. This modified schedule lasted for five years, referred to as the BOND participation period. The treatment group also received work incentives counseling on par with what is available through the Work Incentives Planning and Assistance program under current law. The Stage 2 treatment group was divided into two arms. Both treatment arms received the $1 for $2 benefit offset. Additionally, one treatment arm received enhanced work incentives counseling and the other standard counseling. The goal of Stage 2 was to learn about differences in the effectiveness of passive versus active counseling, as the enhanced counseling involved staff reaching out to beneficiaries unsolicited.

The change to an annual accounting system added an additional layer of administrative complexity. Each year, beneficiaries who intended to work were asked to submit an Annual Earnings Estimate (AEE).\(^{28}\) If anticipated earnings exceeded the BYA, monthly benefits were reduced using the $1 for $2 offset calculation applied to the AEE. If actual annual earnings differed from the AEE, SSA made an adjustment either by paying additional back benefits or recouping overpayments.

The final BOND evaluation report focused on total earnings and total SSDI benefits between 2011 and 2015 as the confirmatory outcomes of interest. Relative to the control group, the Stage 1 (nationally representative) treatment group received average total benefits over the five-year window that exceeded those of the control group by $655 (relative to a control mean of $53,490). The mean earnings of the treatment group were nearly identical to that of the control group.\(^{29}\)

\(^{27}\) See the final evaluation report by Gubits et al. (2018a/b) for additional details on BOND.

\(^{28}\) This process was similar to the one used to apply the Social Security retirement earnings test.

\(^{29}\) Though employment per se was not the major focus of the BOND reports, the Stage 1 treatment group had a statistically significant small increase in employment (one-third of a percentage point impact on the probability of any employment 2011–2015 [control group mean was 22.1 percent] and on the number of years of employment [+0.01 of a year, control group mean was 0.67 years]).
Though mean earnings were similar for the Stage 1 treatment and control groups, the distribution of earnings differed in several notable ways that suggest that BOND had opposite-signed impacts on earnings in different parts of the earnings distribution. A larger fraction of the treatment group than the control group had annual earnings above BYA (+0.23 percentage points from a 2014/2015 base of 2.8 percent). However, a smaller fraction of the treatment group had annual earnings above 200 percent of the BYA (−0.1 percentage points from a 2014/2015 base of 1.4 percent) or above 300 percent of BYA (−0.08 percentage points from a 2014/2015 base of 0.76 percent). The small magnitude of BOND’s impacts on the share with earnings above BYA, 200 percent BYA, and 300 percent BYA in part reflects the small size of the subgroup that was directly affected by BOND, those who complete a TWP. Only 7.1 percent of BOND participants completed a TWP during the first five years of the demonstration, so the program’s impacts are meaningfully larger as a share of that directly affected population.

BOND’s opposite-signed impacts on the share of beneficiaries working above BYA (positive impact) and on the share working above two and three times BYA (negative impacts) illustrate that a given change to the SSDI benefit payment formula can provide different a different incentive regarding the choice of whether or not to work from the choice of how much to work conditional on working. To see this, first consider how the addition of the $1 for $2 benefit offset affects the incentive to work above BYA for an individual who in the absence of the offset would not have worked above BYA. For an individual in this circumstance, BOND provides a substantial increase in the incentive to work and earn above BYA by reducing the high implicit tax on the first dollars earned above BYA. For an individual who not would have worked above BYA in the absence of the offset, work and earnings should tend to increase due to a substitution effect. The increase in the share of the Stage 1 treatment group earnings above BYA relative to the control group is consistent with this prediction.

In contrast, for individuals who would have worked even in the absence of BOND, the benefit offset structure provides an incentive to reduce the number of hours worked. Consider an individual who would have worked with earnings in the offset region above BYA in the absence of BOND. For this individual, the introduction of the BOND offset increases monthly benefits paid if work behavior is held fixed, and the $1 for $2 structure increases the marginal tax rate on earnings from 0 percent to 50 percent. The “windfall” of higher benefits with no change to work behavior should tend to reduce labor supply (hours worked) through an income effect. Additionally, the 50 percent marginal tax rate under BOND compared to zero marginal tax under current rules will tend to reduce labor supply through a substitution effect. These responses tend to lower earnings conditional on working above BYA. The decrease in

---

30 Somewhat similar findings were obtained in the Benefit Offset Pilot Demonstration (see Weathers and Hemmeter 2011).
the share of the Stage 1 treatment group working above two times BYA and above three times BYA relative to the control group is consistent with these predictions.

The results for BOND Stage 2 were similar in many ways to those of Stage 1. Relative to the control group mean of $49,633, the Stage 2 treatment group received average total benefits over the five-year window that exceeded those of the control group by $1,791 for the treatment group that received standard counseling and by $1,997 for the group that received enhanced counseling, both statistically significant at the 5 percent level. The treatment group’s earnings were not statistically significantly different from the control group’s earnings. Both Stage 2 treatment groups were more likely than the control group to have worked at least one year with earnings above the BYA. In contrast to Stage 1, the Stage 2 treatment groups were not statistically significantly less likely to earn above 200 percent BYA and 300 percent relative to the control group, suggesting the income effects generated by BOND had smaller consequences for the recruited volunteer with an expected higher interest in working. Finally, the treatment arm that received enhanced work incentives counseling did not have statistically significantly different benefits or earnings from the first treatment arm that received standard counseling.

A somewhat complicating factor in the interpretation of the results is that members of the BOND Stage 1 and Stage 2 treatment groups had until 2017 to complete the TWP and begin a BOND participation period, so those who entered the period near that end date offset could receive offset payments through 2022. As of the end of 2020, the number of participants still generating payments was very small.

BOND is important because it is the first effort involving a nationally representative sample to address the cash cliff problem that has drawn widespread criticism from the research and advocate communities. The results are not encouraging, however, because smoothing out the cliff can both increase and decrease work effort, and the net effect was approximately zero. BOND also had other difficulties. One was that the 12-month accounting period was difficult to implement, with the consequence that it is not clear whether the beneficiaries experienced the benefit offset in the way it was intended to work (Wood and Goetz Engler, Chapter 9 in this volume). In addition, participants in the treatment groups did not understand the benefit offset very well. However, the hypothesis that the relatively modest impacts of BOND were driven by a lack of understanding of the offset rules is not supported by the finding in Stage 2 of similar earnings impacts for the treatment arm that received

---

31 However, employment effects for Stage 2 were larger than for Stage 1: 2–3 percentage points on the probability of any employment during 2012–2015 (Stage 2 control group mean was 52.5 percent). Also, the significance of the mean earnings impact was sensitive to whether multiple comparisons adjustments were made to the standard errors. When no such adjustment was made, the mean earnings difference was statistically significant at the 5 percent level for the first treatment arm.
enhanced work incentives counseling and the treatment arm that received standard counseling.\textsuperscript{32}

**Promoting Opportunity Demonstration (POD)**

Similar to BOND in several ways, POD was mandated by Congress in the Bipartisan Budget Act of 2015. POD began in 2018 and will conclude in 2021. POD studies the impact of replacing the SSDI cash cliff with a formula that reduces benefits by $1 for every $2 earned above a particular threshold. The POD sample consists of 10,070 volunteers, ages 20–62 over eight sites, who were randomly assigned to three equally sized groups: two treatment groups and a control group that faces the normal SSDI program rules. The volunteers could be current-pay beneficiaries or individuals whose benefits had been suspended for working over the SGA level. They could be concurrent with SSI but could not have a pending work continuing disability review (CDR). The duration of the treatment is from the time of enrollment to the end of the demonstration in 2021.\textsuperscript{33}

The POD intervention differs from that of BOND in several ways. Most importantly, POD eliminates the TWP and EPE.\textsuperscript{34} In addition, the benefit offset is operated on a monthly rather than yearly accounting period, which likely simplifies the program from both the operators’ and participants’ points of view. Combining both of these, from the first month of POD, beneficiaries face a benefit formula that reduces monthly benefits by 50 percent of any monthly earnings beyond a threshold called the TWP threshold, which is lower than the SGA level. If impairment related work expenses (IRWE) exceed the TWP threshold, the offset is applied to earnings above IRWE (with IRWE capped at SGA). For one treatment group (T2), there is an additional rule that a beneficiary loses eligibility for SSDI if earnings for 12 months are high enough to drive benefits to $0. For the other treatment group (T1), eligibility is not contingent on past earnings in this way, and benefits return in any demonstration month where the beneficiary’s earnings are sufficiently low (i.e., the same benefits formula is applied every month). Finally, all participants in POD are volunteers, while BOND Stage 1 involved a larger nationally representative sample of beneficiaries that was not restricted to volunteers.

The Mathematica interim evaluation report (Mamun et al. 2021) reports that through 2019, the first year after enrollment was completed, a higher fraction (24 percent) of those in the treatment group are using the offset than was the case for the

\textsuperscript{32} Difficulty in understanding work incentives rules is a general problem, even in the current-law program.

\textsuperscript{33} See the POD design report by Wittenburg et al. (2018) for additional details.

\textsuperscript{34} In addition, participants were told that their work and earnings during the demonstration would not count against TWP or EPE accumulations after the demonstration ended, and that after the demonstration ended they would return to the same TWP and EPE status they had prior to enrollment.
Stage 2 (recruited participants) treatment group in BOND (14 percent used the offset in at least one of the first three years). This could be for several reasons, the most likely being the elimination of the TWP. But, among other possible explanations, the higher take-up could also be a result of the quicker responsiveness inherent in a monthly accounting period than in an annual one, and an easier-to-understand benefit formula.

The impact estimates from the interim report do not find evidence that POD had an impact in its first year on any of the four primary outcomes of interest (Mamun et al. 2021). During 2019, the treatment groups (pooling T1 and T2) had nearly identical average earnings, probability of substantive employment (earnings above the SGA level), annual SSDI benefits, and total annual income to the control group. The treatment group experienced a substantial increase in employment-related activities such as job seeking relative to the control group, suggesting that impacts on primary outcomes could occur in later years. The report also highlights some implementation challenges. About 6 percent of the treatment group withdrew from the demonstration in the first year, and survey results found that less than half of respondents from the treatment group understood the offset rules.

There are a number of issues with POD that will complicate the interpretation of its findings. One is that some members of the treatment group will be made worse off by POD than they would have been under current-law rules. The clearest case of this is that current rules allow unlimited earnings with no benefit reduction during the TWP, whereas POD reduces those benefits for individuals working over the TWP threshold. In addition, because the TWP threshold is about 75 percent of the SGA threshold, those who would work just below the SGA level under current-law rules will receive a lower benefit under POD, as well. These features interact with the provision that members of the treatment group can always opt to return to current-law rules if they wish. It is possible that those who realize they are worse off under POD than under current law might opt out of the demonstration. This possibility will accurately generalize to a permanent national program that allows people to opt in and opt out in a similar fashion, but not to a program that allows opt in without the option to later opt out.

There is also the overarching question of whether the same opposite-signed effects on work effort that occurred under BOND will also occur under POD. Standard analyses of the different groups involved implies that those opposite-signed effects should indeed occur. Whether they will occur, and at what magnitude, will be revealed by the findings.

One unique aspect of POD is the T1 group, which will be able to work above the TWP threshold for the entire period of the demonstration without jeopardizing eligibility for benefits. Much of the work disincentive in the SSDI program comes from the TWP and EPE structure, which makes termination of benefits an eventually likely outcome for anyone who completes a TWP. That threat is removed for the T1 group during the demonstration, and it will be interesting to see how they respond.
Whether the demonstration period includes a sufficient amount of time to adequately test for this effect is also an open question.

**State Partnership Initiative (SPI)**

The SPI was a collection of state-level demonstrations fielded from 1998 to 2004 intended to draw SSI and SSDI/SSI concurrent beneficiaries into employment service through Workforce Investment Act One-Stop Career Centers. SSA reviewed state-level programs in 2001 and provided funding during 2001–2004. In all, 12 state-level demonstrations (in California, Illinois, Iowa, Minnesota, New Hampshire, New Mexico, New York, North Carolina, Ohio, Oklahoma, Vermont, Wisconsin) received funding under the SPI umbrella. Though many of the major components of SPI involved employment services, we include SPI in this section of our review covering demonstrations focused on financial incentives because an important component of SPI involved modifying financial incentives related to work. (We discuss it in the next subsection.\(^{35}\))

The 12 SPI projects all aimed at increasing access and use of employment services for SSI and concurrent SSI/SSDI beneficiaries who demonstrated an interest in work, and states differed in the combinations of employment services they provided. The SPI projects all recruited SSI-only and concurrent beneficiaries through One-Stop Career Centers, which naturally screened for an interest in working. The demonstrations created by participating states provided different bundles of services aimed at removing employment barriers. In varying combinations across demonstrations, these services included benefits/work incentives counseling, case management, placement assistance, job training services and supports from local mental health and developmental disability service providers, workplace accommodations, job service vouchers, psychosocial rehabilitation, peer mentoring, situational assessment, Medicaid waivers, One-Stop Center services, and outreach to employers.

For the most part, SPI was implemented in a manner that did not involve randomly assigned control groups, which limits the lessons that can be drawn. Only 4 of the 12 programs constructed comparison groups, and only 1 of those 4 used random assignment. Uneven data collection for these comparison groups relative to data collection available for program participants provided additional obstacles to formally evaluating these programs. Of the four programs that defined comparison groups, three of the states’ treatment groups saw larger employment increases than their comparison groups, and one state’s treatment group saw smaller employment increases than its comparison group. For New York, which had an experimental design, Peikes, Moreno, and Orzol (2008) compared the experimental and non-experimental outcomes. They found the two methods gave results that differed in both magnitude and direction.

\(^{35}\) See the SPI *Conclusions* report by Kregel (2006a) for additional details on the SPI demonstration, and see the final evaluation report by Kregel (2006b) for a focus on this waiver component of the broader SPI demonstration.
likely because unobservable characteristics, such as motivation, locus of control, and health status were more important than the observable characteristics that were available for the treatment and comparison groups.

**SPI’s SSI Work Incentives Demonstration Project**

Though SPI focused primarily on providing employment services, 4 of the 12 SPI state programs participated in an additional demonstration aimed to reduce the financial disincentives to work generated by the normal SSI benefit rules. Also known as the “SSI Waiver Demonstration Project,” the SSI Work Incentives Demonstration Project served a recruited subgroup of the participants in the SPI demonstration in the four participating states (CA, NY, VT, WI). The project took place during 2001–2004, beginning later than the main job-services component of SPI. Across the four participating states, programs recruited 1,918 participants. All participants were either SSI recipients or SSI/SSDI concurrent beneficiaries.

The project provided four interventions. First, the waiver provided a “one-for-four” BRR. Normal program rules reduced SSI benefits $1 for every additional $2 earned, after an exemption for the first $65 of monthly earnings. Under the waiver, this 50 percent implicit marginal tax rate on earnings was reduced to 25 percent.

Second, the waiver provided more favorable treatment of unearned income related to work activity, examples of which include Unemployment Insurance, workers’ compensation, and state disability benefits. Normal SSI rules exempted $20 of unearned income and applied 100 percent BRR for any additional unearned income. For waiver participants, this form of income was treated as earned income, subject to the higher $65 exemption and subject to the reduced one-for-four BRR beyond the exemption.

Third, the waiver allowed the accumulation of assets in an “independence account.” Under this provision, participants were allowed savings of up to 50 percent of gross earnings up to $8,000 per year in a checking/savings account that was excluded from the SSI asset test. The exemption lasted until the end of the waiver program (September 30, 2004) and allowed accumulated assets to be spent down over a 24-month period.

Fourth, the waiver suspended medical CDRs in certain cases. Medical CDRs were suspended for SSI-only waiver participants classified as “medical improvement possible” or “medical improvement not expected.”

Unfortunately, the same problem of lack of credible comparison groups that impeded the evaluation of SPI proved even more serious for evaluations of the SSI Work Incentives Demonstration Project. Of the four states, only New York constructed a comparison group (without random assignment), and data sufficient for an evaluation were not collected for the group.
Demonstrations Modifying Vocational Rehabilitation or Other Employment Services

*Project NetWork*

Fielded between 1992 and 1995, the Project NetWork demonstration studied the effects of intensive case management and employment services on return-to-work outcomes. The demonstration recruited volunteers who were SSDI beneficiaries or SSI recipients or had applied for SSI. It randomly assigned eligibility for services from four separate case management models, with each model implemented at two of eight demonstration sites. Recruited participants were randomly assigned to either a treatment group that was eligible for case management services following the model associated with their demonstration site or to a control group. All participants, including those in the treatment group and the control group, received temporary exemptions to several of SSDI’s normal rules regarding the consequences of earning above the SGA level (12 months of work were shielded from the SGA or TWP calculations, and rules were modified regarding CDRs for SSI beneficiaries in Section 1619 status).36

Each of the four case management models included in the demonstration provided more intensive services than were typically provided by state VRs. The models differed primarily by the agency or organization that provided the services. Model 1 offered case management from SSA staff. Model 2 offered case management from private rehabilitation organizations contracted by SSA. Model 3 offered case management provided by state VR agency staff stationed in SSA offices. Finally, model 4 offered less intensive referral management services provided by SSA staff. A total of 4,160 participants assigned to a treatment group were eligible for one of the four models.

The main outcomes of interest were earnings and SSDI/SSI benefit receipt and total benefits. Compared to the control condition, the treatments overall increased service receipt, but had only small, short-lasting effects on earnings and no detectable impact on benefits. The treatment groups experienced a modest increase in earnings relative to the control group in years one and two of the demonstration, but their average earnings were not statistically significantly different from the control group by year three.

*Ticket to Work*

Passed into law in 1999, the Ticket to Work (TTW) program was a reform intended to promote work among SSDI beneficiaries and SSI recipients by introducing a new reimbursement system for employment service providers that was made available both to existing state VR agencies and, marking a change, to private

36 See the evaluation report by Kornfeld et al. (1999) for additional details on Project NetWork.
providers that the program calls Employment Networks (ENs). Though TTW is a national program, its rollout was staggered in two ways that generated quasi-experimental variation in exposure to the program in its early years that facilitated evaluations of the program’s effects.\(^\text{37}\)

The idea of TTW is to compensate employment service providers when their services help move participants back to work and off the benefit rolls. Once the TTW program was fully rolled out, all SSDI beneficiaries and SSI recipients receive a “ticket” that they can assign to a participating employment service provider, either a state VR or a private EN, which enrolls the participant in the organization’s service plan. The VR or EN can choose to be compensated in one of two ways: If the provider chooses outcome-only reimbursement, the provider begins receiving payments from SSA only if and when the participant stops receiving benefits because of work. If the provider chooses milestone-outcome reimbursement, the provider receives a smaller stream of payments if and when the participant stops receiving benefits because of work, but the provider also receives payments as the participant reaches earnings milestones, representing progress toward program exit, while still receiving benefits. VRs have the option of using the traditional cost reimbursement system that was available prior to TTW, and most tickets are assigned under this option. Under cost reimbursement, the VR can be reimbursed for their service costs if the client/beneficiary reaches SGA employment in 9 months within 12 consecutive calendar months. The VR can choose on a case-by-case basis the cost reimbursement option or their selected payment method under the new options (milestone-only or outcome-only).

Early evaluations of TTW found that payments to service providers under the program’s original rules were typically far less than the providers’ cost of providing services (Thornton et al. 2007, ch. IX). Addressing this shortfall, a set of reforms in 2008 modified the rules governing reimbursement to service providers in several ways that removed risk and increased expected payments to providers, primarily by front-loading the reimbursement schedule to generate payments earlier in the return-to-work process. First, the reform shortened the payment period for outcome payments for SSDI clients under both milestone-outcome and outcome-only reimbursement so that ENs could potentially receive full payment within 36 months compared to 60 months under the original regulations. Second, the reform added milestones that triggered payments earlier in the return-to-work process under the milestone-outcome option. Third, payments under the milestone-outcome option were set closer to the larger outcome-only payment amount. Fourth, the reform increased payments for services to SSI-only recipients to be closer to the higher amount for SSDI beneficiaries.

The evaluations of TTW have focused on the three main outcomes related to the program’s original goals; receipt of services, earnings, and benefits paid (Thornton et

\(^{37}\) See the final evaluation report by Thornton et al. (2007) for additional details about Ticket to Work.
The quasi-experimental comparisons made in the TTW evaluation examine in the program’s early years, because the comparisons are generated by the program’s staggered rollout. The rollout was staggered for three groups of states, with Phase 1 states first mailing tickets in February 2002, and Phase 3 states first mailing tickets in November 2003. Mailings were also staggered within groups of states over 10-month windows by last digit of participants’ Social Security number (SSN), providing a second source of variation in exposure to the program across individuals that could be studied for program evaluation.

Two main estimation strategies were used to evaluate the early impacts of TTW. The first evaluations conducted by Mathematica used a difference in differences strategy based on a comparison of individuals in the early-rollout Phase 1 states to individuals in the later-rollout Phase 2 and three states. These evaluations found that TTW increased receipt of services relative to the prior VR-only environment but did not find impacts on earnings or benefits paid (Thornton et al. 2007; Wittenberg et al. 2007). This estimation strategy has both advantages and limitations. The main advantage is that the method estimates the intent to treat impact of the policy on the full population of SSDI beneficiaries. The weakness is that the assumption under which the approach is valid, namely that the Phase 1 states would have exhibited similar trends as Phase 2 and three states in the absence of the policy change, is a strong assumption that is not supported by several placebo tests presented in the evaluation. The difference in differences impact estimates were therefore interpreted cautiously.

A second estimation strategy used by Stapleton, Mamun, and Page (2014) exploited a cleaner source of quasi-experimental policy variation, the staggered mailing of tickets to individuals within states following TTW’s initial rollout. Comparing individuals who were mailed tickets in earlier versus later months based on the exogenous assignment of mailing dates, Stapleton and colleagues arrive at a similar conclusion that TTW caused an increase in service receipt did not have an effect on the number of months with no benefits paid due to work. The main limitation of this evaluation approach is that the impacts that are identified are very short-term effects for the subgroup of individuals who are “compliers” in the sense that their participation decision was influenced by the month that their ticket was mailed, a population who may be affected by the program differently than the full eligible population. The advantage is that the exogenous variation in mail dates allows for credible estimation of this particular causal effect of the policy. Acknowledging these caveats, neither estimation approach finds any evidence that TTW had detectable impacts on exit for SSDI. The 2008 reforms to the reimbursement structure were not rolled out in a way that generated natural comparison groups. However, during 2007–2010 the number of ENs accepting at least one ticket nearly doubled from 818 to 1,600 (Schimmel et al. 2013), suggesting that entry of service providers responds strongly to financial incentives.
Ohio Direct Referral Demonstration (ODRD)

The ODRD is an effort sponsored by SSA and the state of Ohio to test a more aggressive approach to the use of VR services. In the normal operation of the SSDI and SSI programs, a large fraction of beneficiaries and recipients do not make use of those services, possibly because they are not aware of them. ODRD instead makes a direct referral to VR, with the goal of increasing the rate of usage of VR. The population enrolled in the demonstration are ages 18–19 and either are SSDI beneficiaries and SSI recipients who are undergoing a redetermination of eligibility or are applicants to SSDI or SSI. This study population was selected because young individuals could be more likely to benefit from VR services, because those who are undergoing redetermination are necessarily reevaluating their employment options, and because making use of VR services during the application phase could lead to better employment outcomes whether the applicant receives an award or not.  

Enrollment in ODRD began in January 2020 but was paused in March 2020 because of the COVID-19 pandemic. It was restarted remotely in July 2020. The goal was to enroll 750 participants in the demonstration, and for the treatment to be one year in duration, with individuals randomized into either the treatment group (who are referred to VR) or the control group (who are not referred to VR) based on the terminal digit of their SSN. The outcomes of interest are whether the individuals use VR services, whether they remain on SSDI or SSI at a later date, and employment levels. There are no results yet from ODRD, but it is one of several demonstrations that seek to intervene early, or at least at a critical evaluation point, in an individual’s engagement with SSDI or SSI. It will be of interest to see whether early intervention has favorable impacts.

Demonstrations Focused on Employment Services for Those with Mental Impairments

Transitional Employment Training Demonstration (TETD)

The TETD was a randomized control trial (RCT) fielded during 1985–1987 studying the effect of providing transitional employment services to SSI recipients with a diagnosis of intellectual development disorder. The outcomes of interest were employment, earnings, use of disregards, and SSI benefits paid. A set of transitional employment services were provided in 13 different sites around the country. More than 13,000 SSI recipients ages 18–40 were contacted and offered participation in the program. About 5 percent took up the offer. Those who took up the offer were those

---

38 See Wittenburg and Livermore (2020) for a review of interventions for youth. We include this demonstration in our review because it concerns changes in the application process for SSDI and the adult SSI program, which many of those reviewed by Wittenberg and Livermore do not. SSA (2020a) provides additional details on the design of ODRD.
most interested in improved employment outcomes, and they were disproportionately young in age. A total of 745 TETD participants enrolled, most with little recent employment history or history of participation in VR. Participants were randomly assigned to a treatment group \((N=375)\) who received services or to a control group \((N=370)\) who did not.\(^{39}\)

The treatment group received employment services designed around three goals: (1) placement in potentially permanent competitive jobs, (2) provision of specialized on-the-job training to be phased out over time, and (3) postplacement support and services to support job retention. All sites were required to provide the same basic set of services, but the specific service packages and associated costs varied across sites, with average costs estimated to be $5,600 per person, with a site maximum of $14,000 per person. All services were time limited (transitional). The control group members were provided no services but were free to seek employment-related services on their own.

The control group engaged in few employment activities other than sheltered workshops, but about two-thirds of the treatment group were placed in at least initial jobs. A large fraction of these jobs were stabilized into longer-term employment. Many of the jobs were “community” jobs, which were integrated with the regular workforce, and a majority of participants in these jobs had a coach or training program assistance and so were supported in their employment. The specific nature of the impairment appeared to have little effect on these outcomes.

Quantitatively, the TETD services led to an 85 percent average increase in earnings over a three-year period relative to the control group (whose earnings also increased over the demonstration period), with some evidence of fade out of this impact over time.\(^{40}\) Over a two-year period, the treatment group showed 10 percent greater use of employment services, a 1 percent increase in supported employment, and a 1.5 percent decrease in the use of sheltered workshops relative to the control group. Additionally, the treatment group experienced an 18 percentage point gain over the control group mean of 56.7 percent as ever having been employed, an increase of 2.5 months in time employed, and a gain of $1,500 in annual earnings (100 percent increase over the control group mean of about $1,500). These employment impacts were accompanied by only a small decline in SSI benefits because some of the increase in earnings experienced by the treatment group was disregarded rather than being subject to the 50 percent BRR.

The outcomes varied widely across the sites. Whether this was because of differences in service delivery, variation in the local labor market, or some other factor could not be determined. However, an examination of the different sites suggested that the impacts were greater in those that emphasized placing participants in potentially

\(^{39}\) See the final evaluation report by Thornton and Decker (1989) for additional details on TETD.

\(^{40}\) In the first two years, treatment group mean earnings were $1,574 greater than the control group mean of $1,556.
permanent jobs as soon as possible, matching jobs and participants carefully, and being flexible in response to individual participants needs.

TETD is one of the few demonstrations that has targeted SSI rather than SSDI. Its impacts were largely positive, which establishes the potential for employment increases to be possible for SSI recipients. As with many RCTs, the participants were volunteers who were most interested in improving their employment. The large variation across sites, common to many RCTs, is a cause for concern. One issue with TETD is that its findings are 35 years old; whether they would hold with today’s caseload is unknown. In addition, new models of providing services to those with mental impairments have been developed since TETD, such as the Individual Placement and Support model discussed next.

**Mental Health Treatment Study (MHTS)**

The MHTS was a demonstration conducted from 2006 to 2010 to test the impact of a new set of employment services to those on SSDI who had a primary diagnosis of schizophrenia or affective disorder. A central part of the MHTS treatment used an approach to employment services called the Individual Placement and Support (IPS) model, which couples obtaining jobs for participants with disabilities with a suite of associated services. These include employment services, mental health services, benefits counseling, and individualized job supports (participants receive individualized supports as long as they want it in their jobs, and their employers are offered support, as well). The IPS model also aims for participants to obtain jobs in the competitive labor market, to be contrasted with sheltered workshops or supported employment jobs. It seeks to move participants into jobs as quickly as possible after IPS enrollment and without lengthy training. IPS also aims to allow the participants to help direct the course of activities according to their own preferences, rather than strictly following a plan devised solely by the program staff (Frey et al. 2011).

The MHTS demonstration invited a random sample of SSDI beneficiaries ages 18–44 in 23 different study sites. The take-up rate was about 14 percent, with a resulting sample of slightly more than 2,000 enrollees. Enrollees were interpreted as those most interested in employment. They were approximately equally randomized into a treatment group and a control group.

The treatment group received those services included in the IPS model, but also a large additional set of services. These included systematic medication management, integrated behavioral health and employment services, comprehensive health insurance coverage, and nurse coordinator counseling. The medical CDRs were also suspended for three years from the date of enrollment. The treatment group received services for a 24-month period.

The control group received only a manual that listed local and federal resources for persons with mental illness, and their medical CDRs were not suspended. However, they were free to seek any services they might be interested in obtaining.
The impact estimates showed a statistically significant increase in employment, with a 61 percent rate for the treatment group and a 40 percent rate for the control group (Frey et al. 2011). The treatment group had higher earnings than the control group did, although still below the SGA level. Service occupations and sales and office occupations were the most common types of jobs. Predictors of increases in employment were past employment history, health status, and the local unemployment rate. Because earnings were typically below the SGA level, SSDI benefits were not reduced. The treatment also improved participants’ mental health.

The MHTS appears to be a successful model for increasing employment among those with certain types of mental impairments. Some regard the IPS model as a successful model and superior to older-style VR (Bond 1998; Bond, Drake, and Becker 2008), although the evidence from MHTS is not definitive on this because the treatment included more than IPS and was limited to certain impairments. The evidence from MHTS is that the major impact is on employment with earnings below the SGA level, rather than higher earnings, and so exit from SSDI should not be expected from this intervention. It also appears that the MHTS treatment, including not only the IPS services but also the rather large number of additional supports and financial subsidies for health insurance costs, is quite expensive—approximately $7,000 per participant. This could make it infeasible financially if the full SSDI population participated, even only those with these types of impairments.

**Supported Employment Demonstration (SED)**

The SED is an ongoing RCT designed to study the effects of IPS and other services on employment, SSDI and SSI receipt, and mental health of individuals who had applied for SSDI claiming mental impairment but had been denied. The demonstration recruited 3,000 denied applicants ages 18-59 who reported wanting to work or were already working at the time of recruitment. The participants were randomly assigned to one of three groups: two treatment groups and one control group. The control group maintained access to the usual available services and were given an informational handbook. The two treatment groups were provided additional services. The three groups were roughly allocated one-third of the sample each (SSA 2020a).

The first treatment group is receiving what are deemed “basic” services based on the IPS model, as described for MHTS above, but supplemented with behavioral health and employment-related expenses assistance. Care management is also provided to help in coordination.

The second treatment group receives all basic services, plus additional medical services in the form of a nurse care coordinator, systematic medication management, and cost-sharing for medications.

SED recruitment began in November 2018 and will provide services to the treatment group 36 months. The main outcomes of interest include employment, SSDI receipt, SSI receipt, mental health, and quality of life. We expect that the SED final evaluation will include a cost-benefit analysis, so the cost of services is another
important outcome. If successful with job placement, SED could also affect program entry by reducing the probability of a participant appealing the initial SSDI denial or re-applying for benefits at a later date.

SED is of particular interest because it targets those with impairments who have been denied an award. This is presumably a population with less severe disabilities than those who obtain an award, and we know less about the impact of employment services on that population. It constitutes an early intervention, which not only has the chance to assist those who are not yet receiving SSDI but still have major impairments, but it also could delay an appeal to reverse the initial denial (although appeals have important time limits). However, it is also likely to be more challenging to provide services to the study population than those on SSDI already because the SED population is less integrated into the community that provides services to people with disabilities. Connecting them to service agencies with which they might not be familiar, and therefore, for the first time, is likely to be an issue in achieving positive outcomes.

**Demonstrations Focused on Expanded Health Insurance Provision**

*Accelerated Benefits (AB)*

The SSDI program awards Medicare coverage to individuals, but only 24 months after they first become entitled to benefits. Some beneficiaries have health insurance coverage from their previous employer, through spousal coverage or other coverage from the family, from Medicaid if they have low income, or other sources; however, about one-fifth have no coverage at all (Michalopoulos et al. 2011). Yet they often have high medical need and make frequent doctor and hospital visits. The AB demonstration was primarily aimed at studying the impact on health status of health insurance access in the 24-month waiting period (and after it). But employment and return-to-work outcomes were also of interest because any improvements in health resulting from insurance coverage could lead to increases in labor market engagement (Michalopoulos et al. 2011).

In AB, about 2,000 SSDI beneficiaries were enrolled between October 2007 and January 2009 in an experiment that randomly assigned subsidized health insurance to approved SSDI beneficiaries before the end of the usual 24 months. Enrollment took place from 2007 to 2009 and the treatment was for 24 months. Participants in AB were SSDI-only beneficiaries (SSI recipients were excluded because they are covered by Medicaid), ages 18–54, who did not have health insurance coverage at the time of enrollment and who had no less than 18 months remaining in their 24-month waiting period for Medicare. The participants were recruited from the 53 largest Standard Metropolitan Statistical Areas.

Participants were randomly assigned to one of two treatment groups or to a control group. The first treatment group (E1) received health insurance coverage for 24 months, spanning the Medicare waiting period. The plan covered hospital, medical,
and drug claims; use of skilled nursing facilities; home health care; prosthetics, vision, hearing, and dental care; and some out-of-network services. Other than a $12 copayment required for the majority of services provided by the plan, participants were not responsible for any premiums or other costs. However, more expensive care such as emergency room services and inpatient care required higher copayments, and limits were placed on some services such as inpatient care for mental disorders, chemical abuse treatment, and use of skilled nursing facilities. The maximum health care benefit available to a participant during the demonstration was $100,000. At the end of the 24-month period, all AB participants transitioned to Medicare.

The second treatment group (E2) received the same health insurance benefit as E1, plus medical care management services; the Progressive Goal Attainment Program, or PGAP® (a self-paced rehabilitation program to improve functioning); and employment counseling services. About 74 percent of the group took advantage of at least one of the three services. The AB intervention did not make any changes to SSDI’s normal rules regarding the impact of work on eligibility, such as the TWP, EPE, and work CDRs.

Members of the control group received the usual SSDI program benefits and services including its health insurance component. As in the usual program as described above, any health insurance coverage would come from sources outside of SSDI, including Medicaid for beneficiaries who meet the Medicaid means test, coverage from a beneficiary’s previous employer, spousal coverage, or Affordable Care Act coverage from an exchange plan.

There are several publications reporting the impact of the AB intervention (Michalopoulos et al. 2011; Weathers et al. 2010; Weathers and Stegman 2012; Bailey and Weathers 2014; Weathers and Bailey 2014). The main impact estimates showed that AB increased health care use and reduced unmet medical needs relative to the control group (over the course of the demonstration, health insurance coverage increased for the control group, as well). The treatment groups experienced improved health status relative to the control group. There were no measurable short-term effects on mortality. AB also increased participation in TTW, which might be expected to increase employment.

For employment, there were no statistically significant differences between E1 and the control group. There were no statistically significant differences between E1 and E2 in employment in the first year in AB, but there was an increase in employment in the second year for E2 relative to E1. There were no statistically significant differences across groups by the third year, by which point participants from all groups had typically reached Medicare eligibility.

These impacts of E2 relative to E1 imply that the provision of the medical services, rehabilitation services, and employment counseling services had an impact even though health insurance coverage and improvement in health per se did not. The timing of the results suggests that members of the E2 treatment group used the services in the first year, then in the second year benefitted from those services by increased
employment. However, the lack of effects in the third year suggests that there was no long-term impact of the increased employment in the second year on later employment levels.

The AB experiment was expensive. During the first year of AB, the cost per participant for E1 was about $31,000 and for E2 about $34,000.

Strictly regarded from a labor market perspective, the AB demonstration shows the importance of the combination of health insurance coverage and medical and employment services to labor force activity of SSDI beneficiaries. However, the lack of a long-term impact on employment suggests that the provision of that insurance and those services, though improving health and other outcomes, is unlikely to have a major impact on employment while receiving SSDI or on exit rates from SSDI. But AB is among those demonstrations that seek to intervene early in a beneficiary’s time on SSDI. It also has the more general implication that early intervention in general could be a path worth pursuing in future demonstrations (Hollenbeck 2021).

A question about the relevance of the AB demonstration is whether the Affordable Care Act might make its effects different today. The Act provides subsidies for individuals who work and earn enough to put their incomes into the subsidy range. And in states where Medicaid expansion took place, childless individuals are more likely to be covered. This might affect the impact of a health-insurance subsidy program like AB.

LESSONS LEARNED

Our review of the 11 demonstrations in the last section leads us to draw several conclusions. These conclusions will lead us to make a number of recommendations for new demonstrations, which we enumerate and discuss in the next section.

1. Most of the efforts to increase employment, earnings, and labor force engagement of SSDI beneficiaries have been disappointing.

The effort to remove the cash cliff tested in the BOND experiment led to higher employment among Stage 2 participants of 2 to 3 percentage points and only a very small increase for Stage 1 participants (0.36 percentage points). BOND also had a null effect on average earnings of Stage 1 and Stage 2 participants. For the Stage 1 group, earnings increased just above the SGA level and decreased among those at higher earnings. For the Stage 2 treatment group, earnings increased just above the SGA level but experienced only small increases at higher levels, leading to a small overall earnings gain.

These small earnings effects could be the result of opposite-signed incentives at different levels of earnings but also possibly the result of problems in implementation and of a treatment that too many participants did not understand. Efforts to add counseling to participants to help them understand the benefit offset had no effect on
earnings but did increase understanding. The Project NetWork demonstration had only small increases in employment and earnings. The TTW evaluation showed increases in service use but only a small fraction of participants took their tickets to a provider, suggesting that little subsequent impact on employment is likely.

Many of these demonstrations faced difficulties in implementation, and in several cases participants appeared not to understand the work incentive rules well, as we mention below. Technically speaking, we cannot rule out that those problems led to the lack of major employment impacts. However, on the key underlying question of whether the low employment rates of SSDI beneficiaries are the result of low residual work capacity, or the work disincentives in the SSDI rules, we can safely say that we have no evidence thus far that the disincentives are the major problem. Analogously, we can say that these demonstrations do not provide evidence of the existence of a large residual work capacity that is untapped and only not exercised because of work disincentives.

2. The demonstrations for SSDI beneficiaries with mental impairments were an exception and showed more favorable effects.

TETD showed favorable impacts on employment and earnings, as did MHTS. MHTS took place much later in time than TETD and showed that applying the IPS model combined with a large number of additional services could raise employment and earnings. It is important to note that MHTS did not provide evidence on the impact of IPS alone, only on the impact when it is supplemented with substantial additional services. It should also be noted that though MHTS did improve employment, the earnings gains were not sufficient to reach the SGA level, so we should not expect this reform to increase exit from the program.

Another issue with MHTS is that the large number of services provided makes it quite expensive, and it is consequently questionable whether it could be applied to the SSDI caseload at a large scale. Still the concept of trying to get beneficiaries to engage in a large number of services simultaneously because the impacts of the services could complement and reinforce one another—that is, that the impact of the total could be greater than the sum of the impacts of the subcomponents if each had been implemented separately—might apply to other interventions.

3. Where there are favorable SSDI earnings impacts, earnings rarely rise to the SGA level.

Most increases in earnings do not rise to the SGA level, in demonstrations both for beneficiaries with non-mental impairments and for those with mental impairments. This necessarily means that increases are small. Whether this is simply because beneficiaries have too many barriers to work or because they intentionally do not want to work at the SGA level because of the risk that it might lead to termination of benefits.

---

41 See Chapter 8 in this volume for a review of the effect of counseling in SSA demonstrations.
cannot be ascertained. But, again, this general result implies that few additional exits from the SSDI program can be expected from the demonstrations tested thus far. A question for policymakers is how much value they wish to put on increasing employment and labor force attachment even if it does not reduce the caseload and does not reduce SSDI expenditures (indeed, in some cases, it could increase them). Some demonstrations (e.g., MHTS), though not increasing earnings to the SGA level, had favorable impacts on mental health, possibly a result of beneficial effects of working more, so there are other possible outcomes for some of the reforms. Answering this question is outside the scope of our review but is a general question about the goals of reform.

One suggestion that has been made to estimate the maximum work capacity of the SSDI caseload is what is called the “Ultimate Demonstration,” (see Gubits et al. 2019), which would offer recipients the opportunity to receive a lower benefit level in return for no earnings restrictions at all and for an indefinite period of time (but medical CDRs would still be conducted). Beneficiaries would have the option of returning to the standard program if they wished. One issue with the Ultimate Demonstration is that the chance that it would receive congressional approval is probably low, so the expenditure of SSDI trust fund dollars on a reform that is unlikely to be accepted might not be desirable. But a more important problem is that such a reform would, like all reforms that essentially lower the BRR, have ambiguous effects on average work effort because some beneficiaries would be induced to work less, not more, because of the income effects of the reform (namely those who would work above the SGA level in the EPE). We know this to be a possibility because it seems to have occurred in BOND.

4. **There are essentially never increases in exits from these demonstrations and rarely reductions in SSDI expenditures.**

We have already noted this as a corollary of the results showing that earnings rarely rise to the SGA level. In fact, many of the demonstrations increase benefits paid, not decrease them. Again, as just noted, the importance of this conclusion depends on whether the goal of these reforms is to lead beneficiaries to leave SSDI and become self-sufficient, or just to encourage them to engage in work while on SSDI, or what weight each goal is given. Nevertheless, even if the majority of beneficiaries who respond to the demonstrations’ reforms do not leave SSDI or reduce expenditures, it is disappointing that even a small fraction do not leave.

---

42 Another option is not to reduce the benefit at all when offering the program, which could yield a quite different result.

43 Even if not likely to be politically feasible, the Ultimate Demonstration would provide information on the extent of residual work capacity of the beneficiary population.
5. **SSDI financial incentives do not work so well.**

Despite the general objection by many to the SSDI cash cliff, smoothing it out does not appear to increase earnings and could decrease earnings because there are offsetting impacts to the increase in earnings of beneficiaries initially below the cliff. Smoothing the cliff could be desirable for other reasons (such as a more equitable treatment of those just below and those just above the SGA level), but increasing average work effort should not be one of them. Having said this, we must repeat the caveat that we noted in Lesson 1 above that difficulties in implementing and lack of understanding of the rules could have contributed to the lack of effects of smoothing out the cliff. But in this case, there are theoretical reasons, as we have emphasized, that net effects can be small or wrong-signed. Yet another possibility is that even after smoothing out the cliff, the BRR is still too high. We discuss this further in our recommendations for new demonstrations.

6. **Implementation and operational constraints are real.**

Some of these programs (BOND is the clearest example) have treatments that require SSA to make changes in the program that are difficult to implement and would constitute a barrier if they were to be scaled up (Chapter 9). This is not too surprising given the evidence that SSA has difficulty implementing the work incentives rules under current law (Wittenburg et al. 2012, 7) and the reforms in the demonstrations are often even harder to implement. This suggests that the design of future interventions should give simplicity of implementation high priority. At the same time, it should be recognized that the underlying problem is that there is a real tradeoff, because trying to provide the optimal incentives to SSDI beneficiaries and SSI recipients can necessarily lead to complexity. Something might need to be given up to achieve simplicity; it is not costless.

7. **Understanding the treatment is also key.**

Demonstrations that have complex treatments not only are difficult to implement but also are difficult for participants to understand. Treatments that are not understood should not be expected to have much impact. If implementation difficulties occur at the same time, participants are likely to be even more confused about the rules. Again, there is a tradeoff, because designing rules to have optimal work incentives can lead to complexity. So again, something might have to be given up.

As we have noted previously, both implementation difficulties and lack of understanding also hinder the interpretation of the results on program impact, because

---

44 For other examples of implementation difficulties, see GAO (2017) and SSA (2016).
45 POD attempted to do that to some extent.
46 Sometimes the rules create uncertainty, as in the cases of BOND and POD where members of the treatment group may have had concerns about overpayments and having to return money later.
it cannot be known whether the impact is a result of the treatment itself, implementation deficiencies, or lack of understanding. Nevertheless, the question is whether the rules that have been tested in existing demonstrations could be implemented better or if they could be explained to beneficiaries and recipients better than they have been. Though there is no rigorous demonstration evidence on this—because we do not have demonstrations that just vary the effort put into implementation or the effort put into helping participants understand the rules while holding the actual rules fixed—we suggest that it could be that the rules themselves are simply too difficult to implement and understand.

We should also note that there is also a question of whether beneficiaries understand the current-law incentives of the SSDI program (Wittenburg et al. 2012). What this implies is that a demonstration needs to have rules that are clearer to understand than current-law rules, comparatively speaking. Having easier-to-understand rules could be part of the treatment in addition to the other reforms, and the experimental-control difference in outcomes would reflect both.

8. Only a small number of SSDI beneficiaries take up most of the programs.

The take-up rate of the offer of services in the treatment group, as well as the take-up rate of the offer to participate in demonstrations in the first place, is typically quite small. This could be because the treatments being tested are not attractive enough to participants for whatever reason, but it also could reflect that many fewer beneficiaries have substantial residual work capacity than has been thought previously. We suggest that expectations for the fraction of SSDI beneficiaries who have sufficient residual work capacity to take advantage of the programs should be lowered, although recognizing that the size of residual work capacity in the SSDI caseload is still uncertain. Maestas, Mullen, and Strand (2013), for example, have suggested that for those SSDI beneficiaries who had been just on the margin of being accepted into the program (about 12 percent of beneficiaries), residual work capacity is quite high. Of this marginal group, those who were not awarded benefits were 28 percentage points more likely to be employed two years post-determination than otherwise identical applicants who were awarded benefits. However, whether the other 88 percent of beneficiaries have anything close to that level of capacity is questionable.

An exception is BOND Stage 2, which included two treatment arms that differed only in the intensity of work incentives counseling. Those receiving enhanced work incentives counseling had similar employment outcomes to those receiving the standard version, consistent with the idea that the rules themselves likely present a larger barrier than does a lack of counseling effort.

Awards and rejections for the set of applicants involved in this comparison can be thought of as randomly assigned under the study’s identifying assumption that a particular component of the assignment of SSDI applicants to disability examiners is as good as random (Maestas, Mullen, and Strand 2013).

The 28 percent reduction is relative to a small base. The results are sensitive to what earnings range is being considered.
9. Many fewer demonstrations have been tested on the SSI population, but the simple evidence from SSI is not promising.

The SSI program should be of greater interest than it has been because it has a 2-for-1 benefit offset already, it has no SGA rules that create cash cliffs, and it does not terminate recipients from the program after sufficient periods of work over the SGA level, as the SSDI program does.\textsuperscript{50} But despite these much larger incentives to work, only 7 percent of SSI recipients ages 18–64 did so in 2019 (SSA 2020h, Table 42). Almost all SSI recipients who do, work below the SGA level even though they could continue to work above it without loss of eligibility. This suggests again that financial incentives and the SSDI program structure might be less important barriers to work than just the small numbers of recipients with substantial residual work capacity. A qualification to this conclusion is that the SSI program caseload is somewhat different from the SSDI caseload, and because SSI recipients have less work history than SSDI beneficiaries, SSI recipients could have lower residual work capacity to begin with.\textsuperscript{51}

10. Early intervention might have promise.

We reviewed three demonstrations that intervened early in the SSDI application and post-award time period (AB, ODRD, and SED).\textsuperscript{52} Two are ongoing, but the one that has been completed (AB) had favorable employment and earnings impacts from early health insurance provision combined with rehabilitation and employment services during the first 24-month period of SSDI eligibility. Though the effects were not long-lasting, this could be simply because the treatment ended at 24 months, when both treatment and control group members began to have the same health insurance coverage. This suggests that other early interventions might be worth considering for future demonstrations, although clearly AB is only one demonstration, and there have been many more early interventions proposed (for a list of several, see Chapter 5 in this volume).

SUGGESTIONS FOR FUTURE PROGRAM REFORMS AND DEMONSTRATIONS

Here we list a few ideas for future program reforms, based on those lessons learned combined with our general sense of what the barriers are to successful interventions. We divide our ideas into two groups, the first consisting of ideas for

\textsuperscript{50} Note that both the YTD and SPI did test reforms on SSI, however.

\textsuperscript{51} The information in DI & SSI Program Participants: Characteristics & Employment, 2015 (SSA 2020d), especially the section “Work Activity Before and After Award” (pp. 5–11), is useful on the work aptitude of SSI recipients (and SSDI beneficiaries).

\textsuperscript{52} Pre-application interventions include Retaining Employment and Talent after Injury/Illness Network (RETAIN), Promoting Work through Early Interventions Project (PWEIP), and one Youth Transition Demonstration (YTD) site, described in other chapters in this volume.
substantive programmatic changes and the second consisting of ideas for addressing design issues in demonstrations in general.

**Programmatic Changes**

**Earned Income Tax Credit.** There is a vast research literature on the effect of financial incentives on the work incentives of transfer program recipients of various kinds, with the most common topic being the effect of lowering the BRR imposed on increases in recipient earnings. The overall conclusion of that literature, in our view, is that the effects of such a lowering are small if not negative. The results of the BOND experiment are roughly consistent with that literature. Aside from the issue of opposite-signed effects, many analysts also believe that a 50 percent tax rate on earnings is far too high. Such tax rates are rarely present in most countries’ income tax programs, for example, and tax rates that high are mostly considered onerous. Transfer program recipients might be no different from the rest of the population in this regard, and in fact, they might be more conscious of benefit losses of that magnitude if their marginal utility of income is greater than that of higher-income individuals.

The most successful financial reform in other transfer programs in the last few decades has been the Earned Income Tax Credit (EITC). The EITC not only does not tax recipient earnings, it subsidizes them, at least when recipients move from nonwork to work and when increasing earnings at low levels. The research evidence has shown favorable impacts of the EITC on individuals who are initially concentrated at nonwork and whose work decisions are sensitive to financial considerations. These favorable impacts are much greater than the effects of lowering the benefit offset rate in a transfer program, presumably because recipients do not like to have their benefits reduced much at all, but they like being rewarded rather than penalized for working.

The EITC necessarily has a phaseout region, and this could easily have disincentives, as any benefit withdrawal does. However, the evidence showing favorable net impacts on earnings from the EITC has been interpreted as implying that the work disincentives of withdrawal of benefits are not as large as the work incentives of the subsidies, possibly because intensive margin elasticities are smaller than participation elasticities. Some researchers have speculated that this is because the subsidy rate in the phase-in region is high (up to 40 percent of earnings) whereas the BRR in the phaseout region is low (about 20 percent). This suggests that some consideration could be given in a future demonstration to providing high BRRs in SSDI or SSI at low earnings levels and low BRRs thereafter. The question would be whether SSDI beneficiaries, or some fraction of them, have sufficiently high participation elasticities to respond to such a reform.

Gokhale (2013, 2015) has proposed something similar for SSDI, which he calls the generalized benefit offset. However, his plan is quite complex and differs markedly

---

53 In other words, they have high participation elasticities. Single mothers have been identified as one such group. See Hotz and Scholz (2003) and Nichols and Rothstein (2016).
from the EITC. He proposes that workers who earn less than the SGA level have their benefits taxed, leaving them worse off than under current law and discouraging low levels of work. But benefits are extended beyond the SGA level at an increasing rate per dollar of earnings until about three times the SGA level, at which point they are taxed away and end at about four times the SGA level. The benefit schedule is nonlinear, with phase-in and phaseout rates changing for every extra dollar of earnings, which would be difficult for beneficiaries to understand. Gokhale’s phase-in and phaseout rates are also symmetric—both about 36 percent—unlike the EITC, which has a very steep phase-in rate and a very slow phaseout rate.

Many details would have to be worked out in the implementation of an EITC in SSDI. For example, basic questions would have to be answered, such as whether EITC would be subject to the same fixed-length period of the TWP (number of months it takes to complete) followed by an EPE. An alternative would be to restructure the TWP-EPE sequence in some way, or to eliminate them altogether as in POD. Presumably, medical CDRs would continue, but whether work CDRs would continue depends on whether benefit termination policies would continue in the same way they are in the current-law program. Another possibility is to combine the basic EITC with other front-door or back-door policies, as described below.

Reducing the prospect of termination. One of the most likely disincentives to work for SSDI is not the cash cliff but the prospect of benefit termination if the beneficiary works too many months earning at the SGA level. Though this requirement has always been central to the structure of the SSDI program, POD will provide some evidence on the importance of this factor. Whereas one treatment group in that demonstration experiences termination from the program after 12 months of earnings, leading to zero benefits, a second treatment group experiences no termination from the program at any future date (during the demonstration) if working 12 or more such months. If the results show larger work levels in the second treatment group than in the first in those 12 months, it suggests that the prospect of benefit termination is playing a large role in the current SSDI program. However, the impact of the policy is likely to be sensitive to how high earnings have to be to reach the zero-benefits point. If that point is too high, for example, too few beneficiaries will ever reach it, and so this aspect of the reform will not be effective.54

Any policy of this type that alters the point at which termination takes place is often called a “back-door” policy. Aside from altering the number of periods of earnings that result in termination or the level of earnings that count toward termination, an EITC inducement to leave SSDI might be worth considering. Providing an EITC to supplement the earnings of beneficiaries who leave SSDI entirely could provide an inducement for the same reason that the general EITC described above does—namely, instead of just reducing the penalty of leaving the

54 The limited duration of the demonstration may also hinder the effectiveness of the difference in termination rules.
program, it can make the beneficiary financially better off by leaving, relative to current law. Such an EITC would presumably have to be time limited.

**Early intervention reforms.** It is widely believed by labor economists that long periods out of the labor force result in degradation and deterioration of labor market skills and reductions in desire to work and knowledge of working, leading to greater difficulties to return to work later. This suggests that early interventions (“front-door” policies) could have greater chances of success. One example comes from AB, which changed the beneficiary experience in the first 24 months after establishing eligibility. The results from ODRD might yield more information on this approach. SED is rather different, because it targets applicants who have been initially denied eligibility; it is nevertheless related, because it seeks to offer new programs to those initially denied immediately after the decision to deny eligibility has been made.

More generally, the long process that many applicants experience when waiting for decisions and complying with application requirements suggests that more interventions during the application process might be considered. This could involve offering financial incentives to work to those who have not completed an application; or it could involve offering employment services, including counseling, to applicants during the process.\(^{55}\) Though we have not reviewed demonstrations that test such programs, their appeal is that they keep individuals who have experienced a disabling event connected to the world of work and employment during the application process, which might consequently enable them to re-enter into employment more quickly later.

Hollenbeck (Chapter 5) reviews a number of studies proposing reforms at the application stage (including one that offers an EITC to applicants). Many of these proposals offer assistance to applicants who will be denied eventually by regular SSDI procedures, representing an increased value of applying that might draw in other applicants. The proposals also might increase total SSDI expenditures because a new group is being served. However, if applicants who are eventually denied benefits end up having improved outcomes from the services they receive during the application process, that should be regarded as a social benefit. Further, despite this effect, the goal of such programs would be to start offering employment services to applicants who will eventually receive an award. As just noted, the original idea of early interventions is to not wait until much later, in some period after a SSDI award is made, before beginning to reacquaint beneficiaries with the world of work and to establish and maintain work skills and habits.\(^{56}\)

---

55 SED gets at this. Also see “Communicating Employment Supports to Denied Disability Insurance Applicants” (https://oes.gsa.gov/projects/di-denial/).

56 There have also been many reforms suggested for intervening prior to SSDI application, with the RETAIN project being one of these (Hollenbeck in this volume). We do not cover pre-application demonstrations or reforms in this chapter, especially those testing programs on groups that could differ substantially from SSDI applicants.
Simplification and reduction of uncertainty. SSDI rules are fairly complex even in current law, and not all beneficiaries understand them. Moreover, beneficiaries are likely very uncertain as to how the rules will be applied and what the consequences will be for specific employment and earnings decisions they are considering making. The complexity of the work rules is compounded by a cumbersome process by which SSA applies the rules, usually involving an administrative procedure taken only some time after the applicant has worked and after earnings amounts are documented by an often laborious process (Wittenburg et al. 2012). Demonstrations that greatly simplify the rules and reduce uncertainty of the application of those rules are worth considering. Such demonstrations could succeed only if SSA was able, in fact, to make its application of the rules more certain.

An issue with any demonstration on work-rule simplification is the practical problem that SSA has demonstration authority only to test new rules on volunteers, and the question is who would volunteer for a demonstration that was only about rule simplification and nothing else. Further, while using volunteers is acceptable for demonstrations that would never be applied to the full caseload were they made permanent, work-rule simplification would be imposed on the entire caseload (although only affecting those interested in working). In all likelihood, obtaining volunteers for a work-rule simplification reform might be possible only if it were coupled with some more substantive reform in those rules that volunteers would find attractive to try out. Although this would mean that the effects of simplification versus the more substantive reform could not be separated, that would be acceptable if such a condition is also what a permanent reform would look like.

Smoothing the cash cliff. Further demonstrations to test smoothing the cash cliff should be given low priority. The cash cliff could be smoothed to avoid an inequitable bimodal distribution of benefits above and below the cliff, rather than aimed at increasing average employment and earnings, as mentioned above. However, gradual phaseouts of benefits, whether the zero-benefit limit is at the SGA level or something higher, could easily be accomplished as part of other demonstrations that alter the benefit structure in SSDI, without making phaseouts the focus of the demonstration. Gradual phaseout, for example, would be preferable in a demonstration on an EITC.

Time-limited benefits. Some transfer programs offer applicants a time-limited benefit in return for withdrawing their application (e.g., state Temporary Assistance for Needy Families programs have that option, where it is called a diversion policy). Such programs have been suggested for SSDI (Stapleton, Ben-Shalom, and Mann 2019). For example, offering a substantial cash benefit plus VR or employment services for a fixed term might be attractive to applicants who believe they are very likely to return to work. At the end of the fixed term, participants would have the option of again applying for SSDI. Like the early intervention programs discussed earlier, this reform would have the advantage of starting individuals who would eventually become SSDI beneficiaries on the road to labor force reengagement at an early stage. However, part of the intent would be to lead some of those who might
have otherwise gone onto SSDI to instead recover sufficiently that instead of going onto SSDI at the end of their fixed term, they would be capable of attaining self-sufficiency off SSDI.

We noted above the possible attractiveness of early intervention reform, but a time-limited benefit could have a larger application-inducing effect than treatments that only offer employment support services or other non-cash opportunities. Depending on the size of the benefit, the induced application effect could also be quite a bit more expensive. Finally, operating a benefits program parallel to regular SSDI but for applicants is likely to pose more administrative difficulties than just an employment services program. These issues would need to be weighed before considering a time-limited benefit program for SSDI applicants.

**Partial SSDI.** SSDI systems in many countries include some form of partial SSDI provided to individuals whose earning capacity is reduced due to a disability but who still have the capacity to earn at a level above the SGA level. Recent proposals for reforms to the SSDI program in the United States have included calls to introduce a partial SSDI component to SSDI (Maestas 2019). Under this type of program, an individual with a disability whose earning capacity drops (but is above the SGA level) would receive a partial SSDI benefit (proportional to the reduction in their earning capacity) but would be allowed to work while receiving benefits. A partial SSDI program likely would need to be accompanied by more comprehensive initial medical assessments to estimate each applicant’s earning capacity as a specific fraction of their pre-disability earning capacity, as opposed to the current binary classification of disabled versus not disabled (Maestas 2019).

A role for partial SSDI is likely more appropriate now than when SSDI began due to recent trends in the composition of the SSDI applicant pool. Larger shares of SSDI applications now come from individuals with musculoskeletal and mental impairments, more of whom are likely to have a residual capacity to earn above the SGA level than other applicants (SSA 2020b, Table 21). For individuals with residual earning capacity above the SGA level, partial SSDI provides insurance that is tailored appropriately to the size of their loss while providing incentives more aligned with long-term attachment to the labor force compared to current SSDI. Unlike the current SSDI program, partial SSDI would not make an applicant stop working entirely to demonstrate eligibility, and the applicant would face a smaller financial disincentive to work after being awarded.

As with introducing time-limited benefits, expanding SSDI to include partial SSDI for a population that is currently not strictly eligible for SSDI could have application-inducing effects. Partial SSDI also presents a different set of challenges for the design of an informative demonstration. Careful consideration would be required to determine how to structure a version of partial SSDI that is generous enough relative to the current SSDI program to attract demonstration volunteers while still providing evidence that would generalize to the kinds of partial SSDI programs that Congress might consider enacting. One attractive feature of partial SSDI is that it
does not require individuals to fully withdraw from the labor force in order to apply. For this reason, a demonstration might want to recruit volunteers and implement random assignment as early in the application process as possible, even if return-to-work outcomes of those eventually awarded partial SSDI are important outcomes of interest. Another option would be to draw volunteers from other demonstrations that target the application phase or pre-application phase, such as RETAIN. Alternatively, a demonstration could target current beneficiaries who have entered a TWP or who show an improvement to work capacity on a medical CDR, offering re-assessment for partial SSDI as an alternative to the current program rules.

**Demonstration Design Issues**

**Multiple treatment groups.** A single treatment group that receives an intervention with multiple components does not permit attribution of impacts (favorable or unfavorable) to any single component. Some of the demonstrations we reviewed involved evaluations that have two treatment groups, but never more than that. The main issue is the cost of the demonstration if each group is sufficiently large to obtain significant statistical power. One specific case where sufficient power could be retained is where the treatment is scalable; for example, testing various levels of BRRs, various earnings levels at which benefit offset are set to begin, or various accounting periods. By pooling treatment groups with varying levels of the scalable variables and imposing some functional forms on the shape of the response function (for example, by assuming their effect is linear and proportional, as would be the case in linear regression analysis), it is possible for power to be retained in the estimation of that function.

Scalable treatments that can save on statistical power in this way are an example of a type of factorial design, as defined and discussed in Chapters 2 and 3. Also possible is the more general case of factorial designs, where only some combinations of multiple treatments are tested and then the results are used to extrapolate to combinations not tested. We suggest that these and other multiple treatment demonstrations be considered.

Working somewhat against this idea, however, are the results of the MHTS demonstration, which suggest that positive effects may be more likely if multiple treatments are imposed simultaneously. As noted above, SSDI beneficiaries often have multiple barriers to work, and interventions which several components intended to address them all at one time may be more effective than offering each incrementally. Achieving the right balance will require careful consideration of design development.

**Volunteer demonstrations and targeting.** There has been considerable discussion of the advantages and disadvantages of volunteer demonstrations. SSA demonstration authority necessarily requires volunteers and their written informed consent for the participants subject to the intervention (e.g., Stapleton et al. 2020). Some have argued that volunteer demonstrations are biased because they cannot replicate the impact of a permanent program that would impose the intervention on the
entire caseload. Others argue that volunteer demonstrations are superior because they permit estimation of impacts for those for whom the impact is likely to be the greatest, and that such evaluations at least provide an upper bound on general impacts. In their summary of implementation issues in SSA demonstrations, Wood and Goetz Engler (Chapter 9 in this volume) confirm that volunteers for work-incentive demonstrations are indeed those most inclined to work.

These discussions do not always recognize that the question is not what would happen if the entire caseload would take up the newly offered program were it offered to all of them. The question is whether—should the intervention be made permanent and offered to all—the fraction of those taking up the intervention would be the same as the fraction in the demonstration. If those who volunteer for a demonstration are reasonably similar to those who would take up the new intervention were it made national and permanent, then using volunteers should not be a problem.

Our view is based on our conclusion stated in the last section that it is quite likely that only a modest fraction of the SSDI or SSI caseload have enough residual work capacity to ever make major strides in employment, earnings, or program exit. With expectations lowered, concentrating attention on interventions that have effects on that small group of volunteers would seem warranted. However, we propose that interventions that are applied only to volunteers be conducted with the goal not of imposing the features on the entire caseload but only of offering them to a smaller subset of beneficiaries or recipients permanently. There would be no problem in principle of offering employment support programs, for example, which have a fixed number of slots and which are offered to volunteers or to those with identifiable characteristics that suggest higher probabilities of success. Such support programs could be implemented on a national level and on a permanent basis. However, it would be important that any demonstration aim to replicate such an intervention at the national level and be designed with that in mind.57

The idea of having a limited number of programmatic opportunities and encouraging, or even only allowing, those who appear to be most likely to succeed in them to receive them would require identifying those individuals. This raises the larger issue of targeting, which has been used relatively little tested in SSDI demonstrations.58 There is a fairly large amount of evidence from existing demonstrations on heterogeneity in subgroup impacts based on differences in estimated program impacts by age, gender, education, past earnings, or impairment type, for example. The results often show no differential subgroup impacts and hence this body of work is inadequate to guide how targeting might be better accomplished. This is one area where much more research by SSA could be fruitful, with the goal to

57 A rather different problem with volunteer demonstrations is the requirement that the volunteers be offered the opportunity to leave the demonstration and return to the current-law program. Some nationally imposed reforms might permit that, but some would not.

58 See Weathers and Bailey (2014) for a discussion of targeting in the context of the Accelerated Benefit Demonstration.
better identify who is mostly likely to benefit from a reform and to concentrate the reform on that relatively modest number of SSDI beneficiaries and SSI recipients. The work of Maestas, Mullen, and Strand (2013), for example, suggests that residual work capacity can be estimated from outcomes of the application process, suggesting that the evaluations of work capacity arising from that process might help identify those most likely to respond to work incentives reforms. More investigation of the role of specific barriers to work (transportation, specific health barriers, etc.) is also needed; these are rarely included in measures of heterogeneity. The collection of more data on work histories of beneficiaries and recipients would be useful for the same reason. An additional object of research should be identifying work-related correlates of take-up in the first place, which is a necessary precondition to having an impact; heterogeneity in this respect has been little studied, yet the low rates of take-up which motivate this section suggest more research on this issue.

**Estimating more than intent-to-treat impacts.** In line with modern evaluation methodology, most demonstrations estimate only intent-to-treat (ITT) impacts, where evaluations estimate the impact of the offer rather than the impact of actual participation in the intervention. However, ITT impacts can be misleading if large shares of treatment group members do not take up the offered treatment. The low take-up rates in many demonstrations that we noted in the last section make this an issue of some empirical importance. It is likely that policymakers would like to know if a small ITT impact was the result of a small take-up rate but a large response by those who responded, or just a small response by the general beneficiary and recipient population. In the former case, investigation into the types of individuals who responded may also lead to targeting, as discussed above. We suggest that more demonstrations be designed with an aim to estimate additional treatment effects, such as the effect of the treatment on the treated, local average treatment effects, and marginal treatment effects if multiple scalable treatment groups were used, as noted above. As we have already suggested, based on the low take-up rates in the demonstrations we have reviewed and their modest ITT impacts, our view is that only a modest fraction of the SSDI and probably SSI caseloads have sufficient residual work capacity to achieve major increases in employment and earnings. The goal of estimating an average treatment effect on the treated, for example, is precisely the goal of estimating the impact of the treatment on those who take it up and then respond to it. More attention to these alternative treatment effects is consistent with the evidence we have reviewed above on small take-up rates and the likely modest number of beneficiaries and recipients who have major residual work capacity.59

---

59 See Barnow and Greenberg (Chapter 2) and Weathers and Nichols (Chapter 3) for more detailed discussion. In some cases, defining who takes up the program is itself a difficult data and measurement problem because it requires comparing responses of members of the treatment group to members in the control group and defining what take-up and participation actually mean. But if the demonstration designs were to plan this in advance, data and measurement issues might be addressable at the design stage.
**Assessing the effects of limited demonstration duration.** An issue that we find to be underdiscussed in the evaluation reports we have read is the issue of limited duration of the demonstration. Participants know that the treatment is offered for only a finite period of time and that, if they do not exit SSDI or SSI before the demonstration ends, they will revert to the current-law program. Participants are quite likely to be affected by this knowledge and their response to the demonstration might be different than it would be if the intervention were implemented nationally on a permanent basis. This issue was considered in the Negative Income Tax experiments of the 1960s and 1970s, where theoretical analyses showed that the bias in an experiment of limited duration was ambiguous in sign, and could be favorable or unfavorable (Metcalf 1973). However, it is quite likely in the SSA demonstrations we have reviewed (e.g., POD) that the limited duration of the demonstration likely led to an underestimate of the response that would obtain in a permanent program, just because the demonstrations were not long enough for beneficiaries to understand and react to them.  

**Entry effects.** Entry effects of programmatic reforms in SSDI have been discussed for many years. Most have foundered on the problem that entry effects are very difficult to evaluate by RCT methods (Weathers and Nichols, Chapter 3 in this volume). However, it is also clear that the population of individuals considering applying are making decisions that are partly based on the relative financial and other attractions of applying versus not applying. If nothing else, this is demonstrated by the fact that SSDI applications are sensitive to the business cycle and job availability. Given the inability to estimate entry effects credibly with RCT methods, we recommend that non-experimental designs be constructed to estimate entry effects, even if only estimating a range of possible such effects (see Chapter 3 for examples). These could then be combined with the estimated RCT effects of a programmatic reform on outcomes conditional on entry.

**SUMMARY**

In this chapter we have reviewed a number of demonstrations conducted by SSA on SSDI and SSI programs with an aim to increase employment and earnings among beneficiaries and recipients, and to increase exits from the program. Our review of 11 major demonstrations and other tests of SSDI and SSI programmatic reforms leads us to draw several lessons. One lesson we draw is that the impact of most demonstrations

---

60 In one of the Negative Income Tax experiments, three different treatment groups were enrolled, each with a different duration of the offered treatment (SRI International 1983). The difference in response was then used to extrapolate to a permanent program.

61 See Moffitt (1992a) and Hoynes and Moffitt (1999). Entry effects are unlikely to be important for small changes in SSDI work incentives rules, especially if they are poorly understood. But large changes in the experience of a large number of beneficiaries of the program could have a reputational effect that could affect entry.
on the already low rates of employment and exit are disappointingly small. Something of an exception occurs for some tests of reforms in the treatment of SSDI beneficiaries with mental impairments, although those programs are also quite expensive. We also reviewed one demonstration (AB) that intervened at the newly awarded stage that also had favorable effects but was quite expensive. But we also find that for most demonstrations, only a small number of individuals take up work programs offered in SSDI. This includes demonstrations with attempts to improve financial incentives, including the elimination of the cash cliff.\textsuperscript{62} Implementation issues are important as are complexities in the interventions that are difficult for beneficiaries to understand.

Based on these conclusions, we suggest a number of ideas for future demonstrations worthy of further consideration. These include providing even stronger financial incentives to work through EITC-style programs that supplement earnings with benefits rather than reducing benefits; more testing of ways to reduce the work disincentives arising from the prospects of termination from the program; early intervention reforms such as those offering employment services to applicants at some stage of the application process; demonstrations to reduce work incentive rule complexity (either with or without offset); and partial SSDI benefit programs. We also suggest ideas for the demonstration design, such as possible increased use of scalable multiple treatment groups and factorial designs, more intentional planning of volunteer effects in demonstrations in conjunction with determining the effects of targeting, going beyond ITT effects in estimating the impact of program reforms, addressing the problem of limited demonstration duration, and incorporating entry effects.

**ACKNOWLEDGMENTS**

The authors would like to thank Daniel Gubits for important assistance throughout the preparation of this chapter, Jeffrey Hemmeter for comments on an initial draft, and several SSA staff for comments on this draft. We also thank David Salkever and David Stapleton for helpful discussions.

\textsuperscript{62} However, final results from the POD demonstration are not yet available and will be relevant.
Contributors

**Jesse Gregory**, Associate Professor, Department of Economics, University of Wisconsin-Madison—Dr. Gregory conducts research in labor, public, and urban economics. His research often focuses on questions related to the design of policies whose rules may affect individuals’ location choices, including disaster relief policy, low-income housing policies, and place-based hiring subsidies.

**Robert A. Moffitt**, Krieger-Eisenhower Professor of Economics, Johns Hopkins University and Bloomberg School of Public Health—Dr. Moffitt’s research is on the economics of poverty and welfare programs and the economics of the labor market relating to the low-income population in the United States.
Return to Work

Chapter 4

Comment

Hilary Hoynes

*University of California, Berkeley*

The chapter by Jesse Gregory and Robert Moffitt (“The Return to Work in Disability Programs”) provides an excellent and insightful review of the literature on policies to promote returning to work among Social Security Income (SSI) recipients and Social Security Disability Insurance (SSDI) beneficiaries. This wide-ranging review includes a discussion of the conceptual issues around work supports and work incentives in this population, a detailed summary of SSA demonstrations to promote work, and suggestions for future reforms and demonstrations. They conclude that the results are disappointing; that is, there is little evidence that these policies have led to meaningful increases in employment and earnings.

In my comment, I would like to put the chapter and results in some context. I will focus on three main points: (1) why are we focused on work for this population? (2) an analysis of recent trends (pre-COVID) suggests some change is occurring; and (3) thoughts moving forward.

WHY ARE WE FOCUSED ON WORK IN THIS POPULATION?

It is useful to start with some context around why there is so much policy attention on increasing work among SSDI beneficiaries and SSI recipients.

First, as mentioned by the authors, the reforms and demonstrations that they analyze have taken place during a time of large growth in SSDI caseloads and costs. As shown in Exhibit 4.2, the number of workers receiving SSDI increased from 3 million in 1990 to almost 9 million by 2010. Total program costs experienced similar increases during that time. As caseloads have increased, the composition of the SSDI caseload has changed. The share of the caseload with musculoskeletal diagnosis has increased from 20.6 percent in 1996 to 33.6 percent in 2019 (SSA 2020b).

Second, over this period there has been a steady decline in male labor force participation and more recently also a decline in female labor force participation. For example, the percentage of men age 16 and older in the labor force participation fell from 80 percent in 1970 to 75 percent in 1990 to below 70 percent in 2019. After rising for most of the 20th century, the percentage of women age 16 and older in the labor force peaked in the late 1990s and has steadily decreased since.63 There is a large literature that explores these trends to identify the sources of the decline (see recent review by Abraham and Kearney 2020). Disability benefits seem to be part of the story, but there are other factors at play.

---

63 See Figure 1 in Nunn, Parsons, and Shambaugh (2019) for recent data and discussion of trends in labor force participation.
Taken together, the rising costs of the program alongside the declines in labor force participation naturally have led to policy discussions around how to increase employment among SSDI beneficiaries, with the goals of increasing aggregate labor in the economy and (possibly) reducing the costs of disability benefits.

BUT MORE RECENTLY THE TRENDS LOOK DIFFERENT

The discussion above relates to trends occurring over the past two decades or more. The most recent data (pre-COVID pandemic), however, show some important changes in trends. Exhibit 4.3 plots the number of beneficiaries (right axis) and awards (left axis) in SSDI between 2000 and 2019. New awards peaked in 2010 at just over one million and fell steeply to about 750,000 in 2015 before leveling off. Relatedly, and this has received less attention, the employment rate among disabled individuals is on an upswing after a many decade decline (see Exhibit 4.4, from Maestas 2019).

Maestas, Mullen, and Strand (forthcoming) find that the Great Recession, and the changes in labor market conditions, explain some of these recent trends in SSDI caseloads. However, changes in appellate-level decisions may also have played a role. Exhibit 4.5 presents trends in SSDI allowance rates at the hearing level between 1995 and 2018. Allowance rates have declined substantially in recent years—from almost 75 percent in 2005 to 50 percent in 2014. In ongoing work with Nicole Maestas and Alexi Strand, we examine the factors and policies that explain the reduction in allowance rates—with an interest in examining the effects on employment for persons with disabilities.
Exhibit 4.3. Number of Beneficiaries and Awards in SSDI, 2000–2019


Exhibit 4.4. Employment Rate of People with Disabilities

Source: Data from Maestas (2019).
THOUGHTS GOING FORWARD

First, it is important to put the goal of increasing employment and earnings among disabled works into some perspective. The standard public finance framing for considering the “optimal” design of transfer programs is to examine the tradeoffs of protection versus distortion. For SSDI, the goal is to provide protection against disability-related earnings losses but balanced against not inducing labor force nonparticipation among people who could otherwise work. To make sense of the results here (policies and demonstrations around work and work disincentives) we need to also know about protection.

We need more evidence on the effects of SSDI (and SSI) on short- and long-term health, economic and financial well-being (e.g., Deshpande [2016a] on SSI).

Second, we need to build on the encouraging evidence reviewed by Gregory and Moffitt showing that early interventions and health care coverage may increase work among the disabled. This suggests we need to identify ways to keep more SSDI applicants in the labor force—rather than try to affect their behavior once they are receiving disability program benefits. A multi-tiered system might be a good structure. One tier would target those with capacity to work (partial insurance, as promoted by
Maestas [2019]). Another tier would provide health care and short-term income supplements. The third tier would target those with long-term disability without capacity to work.

**Hilary Hoynes**, Professor of Public Policy and Economics, Goldman School of Public Policy, University of California, Berkeley—In addition to her appointment at GSPP, Dr. Hoynes holds the Haas Distinguished Chair in Economic Disparities and co-directs the Berkeley Opportunity Lab. Her research focuses on poverty, inequality, food and nutrition programs, and the impacts of government tax and transfer programs on low-income families.
In their chapter (“The Return to Work in Disability Programs”), Gregory and Moffitt provide an excellent overview of several decades of the Social Security Administration’s (SSA’s) return to work demonstrations. Their findings may surprise and discourage some readers who expected these experiments to increase work for a substantial number of beneficiaries and recipients—especially if they were hoping increased work would increase program exits and thereby reduce costs. For example, Gregory and Moffitt find that most SSA work demonstrations do not meaningfully increase employment, earnings, or labor force participation. Even when beneficiaries’ and recipients’ earnings increase, they rarely rise above the Substantial Gainful Activity (SGA) level. What is more, there are “essentially never” increases in program exits due to work from demonstrations, and “rarely” reductions in Social Security Disability Insurance (SSDI) program expenditures. Only a small number of SSDI beneficiaries try the interventions offered in work demonstrations—possibly, the authors posit, because few have residual work capacity.

Given the very strict medical and vocational criteria for SSDI and Supplemental Security Income (SSI) program benefits, should the fact that few beneficiaries have residual work capacity surprise readers? Consider how the Social Security Act defines a qualifying disability:

The term “disability” means inability to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months. (42 USC § 423(d)(1)(A))

Furthermore, it must be

of such severity that he is not only unable to do his previous work but cannot, considering his age, education, and work experience, engage in any other kind of substantial gainful work which exists in the national economy, regardless of whether such work exists in the immediate area in which he lives, or whether a specific job vacancy exists for him, or whether he would be hired if he applied for work. (42 USC § 423(d)(2)(A))

Perhaps the fact that few SSDI beneficiaries and SSI recipients are able to work at SGA over a sustained period is not evidence that the demonstrations have failed,
but evidence that the disability determination process has succeeded in identifying applicants who meet these extraordinarily strict criteria.

In fact, in addition to their work-limiting disabilities, SSDI beneficiaries and SSI recipients face many other barriers to work.

- Disability beneficiaries have lower education levels than average. Disability can limit a person’s education, and limited education is also correlated with higher levels of disability. Limited education constrains employment opportunities—especially for workers whose disabilities prevent physical work.
- Disability beneficiaries are older than average. The older a person gets, the more likely he or she will acquire a disability. Advanced age poses another employment barrier; it is more difficult for older workers to find jobs or to shift to other kinds of work.
- Disability beneficiaries typically have substantial time out of the labor force. Most SSDI applicants have been out of work at least a year before they even apply for benefits, and for SSI disability, work history tends to be even more limited. The application process for both programs takes months—or sometimes years, if appealed. During this time, workers’ skills and connections atrophy, and gaps in employment history pose an additional employment barrier.
- Finally, people with disabilities face structural barriers to employment. Disability discrimination is pervasive. Many employers do not offer sufficient accommodations to allow disabled employees to work. Transportation can pose another barrier; public transit is often unavailable, inaccessible, or unreliable. And many would-be workers with disabilities do not have access to the health care they need, particularly the long-term services and supports that would allow for employment.

Given that all disability beneficiaries have work-limiting disabilities and most face additional barriers, it is little wonder few SSDI beneficiaries or SSI recipients are able to perform substantial, sustained work, even with incentives to do so.

A DIFFERENT WAY OF MEASURING WORK SUCCESS FOR DISABILITY BENEFICIARIES

If sustained, substantial work is not possible for most disability beneficiaries, is it possible to construct a successful return to work program? The answer to that question depends on how one measures success.

If policymakers expect most disability beneficiaries to return to work, success isn’t likely, past demonstrations show. Instead, policymakers could aim to reduce barriers for those with the capacity and desire to work. Gregory and Moffitt suggest that SSA focus employment interventions on volunteers who seem most likely to succeed. This would allow the agency to tailor interventions to the needs of SSDI
beneficiaries and SSI recipients with specific characteristics and improve their odds of success. Evidence supports this approach, as Gregory and Moffitt point out—SSA’s mental impairment-based demonstrations resulted in rare employment and earnings increases.

If policymakers aim to reduce SSDI and SSI program spending, success isn’t likely, either, based on the evidence. Instead, policymakers could aim to improve beneficiaries’ and recipients’ well-being overall. Many people with disabilities who work report benefits beyond their wages—improvements in mental health, community integration, and a sense of purpose and connection. But these effects are often not measured in demonstration projects. When they are—for example, the Mental Health Treatment Study measured improvement in mental health—they show work can have positive effects beyond earnings. The ongoing Supported Employment Demonstration is another example of a study that measures mental health and quality of life, in addition to financial outcomes.

Shifting from broad-based interventions aimed at reducing costs toward focused interventions aimed at improving quality of life would require not only different services and metrics, but also likely more money, as Gregory and Moffitt point out. If policymakers’ goal is truly to encourage work among disability beneficiaries, it would be money well spent. But if the goal is to save money by reducing enrollment, it seems likely that another round of work demonstrations will only bring more surprise and disappointment.

Volume References


References


References


