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

Till von Wachter. 2021. "An Overview of Current Results and New Methods for Estimating Heterogeneous Program Impacts." In *Lessons from SSA Demonstrations for Disability Policy and Future Research,* edited by Austin Nichols, Jeffrey Hemmeter, and Debra Goetz Engler, 271–322. Rockville, MD: Abt Press.

Howard H. Goldman. 2021. "Comment" (on Chapter 7: "An Overview of Current Results and New Methods for Estimating Heterogeneous Program Impacts"). In *Lessons from SSA Demonstrations for Disability Policy and Future Research,* edited by Nichols, Austin, Jeffrey Hemmeter, and Debra Goetz Engler, 316–318. Rockville, MD: Abt Press.

Nick Hart. 2021. "Comment" (on Chapter 7: "An Overview of Current Results and New Methods for Estimating Heterogeneous Program Impacts"). In *Lessons from SSA Demonstrations for Disability Policy and Future Research,* edited by Nichols, Austin, Jeffrey Hemmeter, and Debra Goetz Engler, 319–321. Rockville, MD: Abt Press.

An Overview of Current Results and New Methods for Estimating Heterogeneous Program Impacts

Till von Wachter

University of California Los Angeles
National Bureau of Economic Research

The numbers of beneficiaries of the Social Security Disability Insurance (SSDI) and recipients in the Supplemental Security Income (SSI) programs grew rapidly over the past decades.[1] At the same time, the demographic characteristics and impairment types of beneficiaries have evolved (Duggan and Imberman 2009; Duggan, Kearney, and Rennane 2015). These trends have raised the question whether beneficiaries and recipients might have greater potential to work now compared to in the past (e.g., Autor and Duggan 2006; von Wachter, Song, and Manchester 2011). To study how to best encourage and support employment among potentially able SSDI beneficiaries and SSI recipients, the Social Security Administration (SSA) has engaged in a series of demonstrations aimed at establishing the effect of various policy changes, incentives, and supports for SSDI beneficiaries' and SSI recipients' employment.

This chapter discusses to what extent the effect of the various interventions tested varies across subgroups of SSDI beneficiaries and SSI recipients. This is an important question because the SSDI and SSI programs insure and serve a broad population of individuals. Current beneficiaries and recipients not only vary substantially in their education, occupation, and skill backgrounds, but they also vary in age, gender, types of impairment, and time spent in the program. Trends that have raised the number of beneficiaries and recipients who are younger and/or are more likely to have impairments associated with musculoskeletal or mental health conditions have further increased that diversity. These are all factors that potentially affect their ability to work and to find work, as well as their likelihood of sustained success in the labor market.

SSA has pursued several demonstrations that aim to provide insights on a range of questions regarding key subgroups of the SSDI and SSI populations. How different beneficiaries and recipients respond to treatments tested in a demonstration is important for several reasons. Documenting the range of possible responses to treatments is helpful for better predicting the potential impact of a tested intervention

---

[1] This increase reversed in 2013 or 2014 and the subsequent decline in participation may have different causes and policy responses. However, the demonstrations reviewed in this chapter are largely a response to the increase in participation, and the lessons from them apply mainly to that situation.

if it was to be offered to the full population of beneficiaries nationwide. Insights on variation in the effects of treatment for certain groups can help in better implementing interventions by informing which beneficiaries and recipients might be particularly responsive to new features of a program and for which the intervention could be further improved.

Evaluation research often considers the nature of treatment effect heterogeneity, such as impacts for subgroups (e.g., see Brock, Weiss, and Bloom 2013; Bell and Peck 2016b; Rothstein and von Wachter 2017). In practice, however, it is often difficult to estimate subgroup effects because of insufficient statistical power, usually due to having smaller sample sizes for subgroups relative to an evaluation's full sample. This limitation means that the role that such differential estimates can play—for example, in better targeting new interventions to particular beneficiaries—is often also limited. A growing literature on estimating heterogeneous treatment effects implies lessons for the next generation of SSA demonstrations. A data-rich environment (such as with some of the SSA demonstrations) combined with analytical/methodological developments suggests some particularly promising opportunities.

This chapter begins with a review of current evidence on the employment potential of SSDI beneficiaries and SSI recipients, with particular focus on variation across subgroups. Then it summarizes evidence of subgroup impact variation among recent SSA demonstrations testing interventions aimed at raising labor force participation and self-sufficiency. For SSDI and concurrent SSDI beneficiaries and SSI recipients, the chapter discusses estimates from the Benefit Offset National Demonstration (BOND) and its predecessor, the Benefit Offset Pilot Demonstration (BOPD); the Mental Health Treatment Study (MHTS); Project NetWork; the Accelerated Benefits (AB) demonstration; and the Promoting Opportunity Demonstration (POD). For SSI recipients, the chapter discusses the Transitional Employment Training Demonstration (TETD) and the Structured Training and Employment Transitional Services (STETS) demonstration.[2] Next the chapter reviews some recent methodological literature on estimating heterogeneous impacts and suggests lessons for future demonstrations. The final section draws some broad conclusions.

BACKGROUND ON VARIATION IN EMPLOYMENT POTENTIAL

Substantial research has analyzed how employment and earnings vary among individuals. Individuals' ability, capacity, and desire to work is sometimes referred to as their employment or earnings or work "potential." A range of factors typically influences such potential. For example, employment potential relates to individuals' ability and desire to work, which is influenced by their health and disability, innate capacity, and preferences, education, work experience, training, family status, child

---

[2]   See Chapter 6 in this volume for more detail on SSI demonstrations.

care, and transportation. The institutional environment—such as taxes and the availability and value of public assistance and transfers—is also a factor.

Employment potential is likely also to be directly affected by SSDI or SSI program design. This is because, according to program rules, an excess level of earnings over some point triggers gradual removal from the program; moreover, receipt of benefits could reduce individuals' need to work.[3] This makes an empirical analysis of employment potential among SSDI beneficiaries and SSI recipients particularly difficult, as it involves an assessment of an inherently unobservable outcome: *What would the employment of a participant be in the absence of program benefits?*

Past research has aimed to estimate the employment potential of non-working SSDI beneficiaries and SSI recipients. Such estimates can provide an indication as to which individuals might be most responsive to inducements to return to the labor force, and at what level one should expect their employment or earnings to be. Individuals with higher work potential are likely to face lower barriers to employment and could be more responsive to financial inducements to return to work. In addition, information on beneficiaries' and recipients' work potential—or factors correlated with greater work potential—could help predict differential responses to non-monetary inducements or other supports to return to work, as well. Responses to monetary or non-monetary inducements to work can also vary across beneficiaries for reasons other than their employment and earnings potential. As discussed later in this chapter, reasons could include variation in how they understand the program or variation in how the program is implemented across time and across space.

Bound (1989) concludes that employment potential of the average beneficiary is small. Considering major impairment groups, age, and gender, von Wachter, Song, and Manchester (2011) find important variation in the employment rates of rejected applicants (which is, admittedly, only partially useful for understanding the employment potential of beneficiaries). Further, they find younger workers and workers with impairments related to the musculoskeletal system or to mental health have higher employment rates than do older individuals or those with impairments of the respiratory or circulatory system. Based on analyzing employment of rejected applicants, Maestas, Mullen, and Strand (2013) and French and Song (2014) also find

---

[3]   The effect of the presence of SSDI or SSI benefits on the labor supply of beneficiaries is sometimes referred to as the "disincentive" effect. This includes a "substitution" effect (that arises because individuals would lose earnings if they were to work more, and hence they work less) and an "income" effect (that arises because individuals would like to work less but cannot because of low income; this effect arises even if individuals were to keep their SSDI or SSI benefits if they work above the SGA level). Strictly speaking, only the substitution effect is considered a program distortion. Though there are a few studies trying to isolate the substitution and income effects, recent research points to an important role of the latter (e.g., Gelber, Moore, and Strand 2017). This distinction, not further discussed in this chapter, could be relevant in its own right for program changes aimed at increasing work potential.

variation in employment potential by age and impairment type. In addition, they indicate some variation in estimated employment potential by prior income, with high-income individuals exhibiting lower employment.

Hemmeter and Bailey (2016) find that in the years after exiting the program, SSDI beneficiaries whose benefits were terminated due to a medical review had a relatively high incidence of earnings (i.e., any employment in a given year), but low rates of consecutive years employed and very low earnings. I suggest this finding reflects an upper bound for labor market outcomes of SSDI beneficiaries because it is for beneficiaries who are perhaps better off health-wise than the average. Hemmeter and Bailey (2016) also find that earnings and employment of those exiting the program can vary substantially with age, impairment type, and time spent on the program. Although some of the characteristics analyzed are correlated (i.e., younger individuals are more likely to have shorter program duration), the results offer some insights regarding variation in employment potential for some groups.

For example, individuals terminated with less than two years in the program are the highest-earning group considered and have about $18,000 annual earnings, whereas those with six or more years in the program have about $11,000 annual earnings. This difference is unlikely explained solely by age. Even among the highest-earning group, only 50 percent of individuals studied had earnings in the five years after program exit, indicating that even those SSDI beneficiaries with employment potential can face substantial labor market barriers and be at risk of financial hardship absent benefits.

Although studied less extensively than SSDI, participation in the SSI program can also affect the future employment prospects of recipients. For example, Deshpande (2016b) finds that children removed from SSI due to age 18 redeterminations recover only one-third of lost SSI cash income, and those who stay off SSI earn only $4,400 on average per year in adulthood. Davies, Rupp, and Wittenburg's (2009) descriptive analysis of human capital development among youth receiving SSI illustrates the heterogeneity among this population and emphasizes the importance of coordination, both contemporaneous and longitudinal, of programs and interventions aimed at supporting these youth.

Overall, employment potential among SSDI beneficiaries and SSI recipients varies in predictable fashion, among others with age and impairment type. However, the groups studied in the literature are quite coarse and far from what would be needed to meaningfully identify specific groups of individuals that could or should be targeted for employment incentives or services. Moreover, the role of different personal characteristics is typically studied separately, but the intersection is likely to be particularly informative about an individual's employment potential. The amount of heterogeneity documented in the demonstration reports provides a sense in which the likely opportunities and needs are likely to substantially differ among beneficiaries. A young beneficiary with an impairment related to mental health will likely have different needs than will an older beneficiary with an impairment of the

musculoskeletal system. Their needs are likely to differ further by years of education, profession, and labor market experience.

Trends in characteristics of SSDI beneficiaries and SSI recipients have tended to further increase the diversity in the characteristics of individuals that SSA serves. For example, the number of SSDI beneficiaries who are younger and have impairments associated with musculoskeletal or mental health conditions has increased in the 1990s and early 2000s (Duggan and Imberman 2009). Between 1988 and 2013, the share of SSI recipients who are younger than age 64 increased by 20 percentage points. The share with intellectual and mental health disorders was 57 percent of the SSI caseload for the working-age population in 2013 (Duggan, Kearney, and Rennane 2015). Because beneficiaries and recipients who are younger and have intellectual and mental health impairments are typically found to have higher employment potential, these changes likely increased the overall employment potential among SSI and SSDI recipients.[4] Whether these trends will continue is a matter of ongoing analysis. Although some researchers have warned the demographic trends may lead to unsustainable increases in SSDI caseloads over the long term (e.g., Autor and Duggan 2006), others suggest that these changes could be temporary, related to the aging of the baby boom generation (Congressional Budget Office 2012; Board of Trustees 2014) and to the increasing share of women in the labor force (Goss 2013). In either case, ongoing changes in the population and in the labor market—such as those brought by the COVID-19 pandemic—and in the SSDI and SSI programs will likely continue to affect the distribution of beneficiaries and recipients and with it the variation in employment potential.

DISCUSSION OF HETEROGENEITY IN ESTIMATES FOR DEMONSTRATION OUTCOMES

SSDI provides income to insured individuals who are unable to engage in substantial gainful activity (SGA) due to a medically determinable physical or mental impairment. SGA occurs if earnings exceed a monthly threshold (SSA 2020e).[5] The SSI program provides income to disabled individuals with limited economic resources, regardless of whether they qualify for SSDI based on their work history, or any individual age 65 and older with limited economic resources.

In addition to providing income support, SSA's programs aim to support the efforts of SSDI beneficiaries and SSI recipients of working age who desire to return

---

[4]   However, Howard Goldman (Comment in this volume) in his helpful discussion raises the important point that for impairments related to mental health, the ability to sustain employment may vary over time as mental health conditions wax and wane.

[5]   Individuals are eligible to receive SSDI benefits ("insured") if they have sufficient quarters of covered earnings (e.g., see https://www.ssa.gov/pubs/EN-05-10072.pdf). For employees, SGA means if working and making more than $1,310 per month in 2021 (or $2,190 for beneficiaries who are legally blind). For self-employed individuals, any month during which work exceeds 80 hours is considered a Trial Work Period month.

to work. SSA's programs do this through a number of work incentives policies and complementary programs providing counseling, among other services and supports. For example, the Trial Work Period allows SSDI beneficiaries to have a total of nine months in which earnings can exceed the SGA level (not necessarily consecutively) over a five-year period. If a worker completes the TWP, then they begin the Extended Period of Eligibility, which is a 36-month period during which the beneficiary is eligible to receive SSDI benefits if earnings drop below the SGA level in a given month.[6] These and other policies are designed to support beneficiaries' return to the labor force (SSA 2020).

Within this broad framework, SSA has implemented demonstrations to test how it might further support SSDI beneficiaries' and SSI recipients' return to work. The interventions studied in these demonstrations test a range of employment inducements and supports. These include, among others, monetary incentives to work above the SGA limit (as in BOND, BOPD, and POD), training (as in STETS), case management (as in MHTS and Project NetWork), job search assistance (as in STETS and TETD), and access to health care (AB). In several cases, demonstrations combine multiple treatments. The two subsections that follow provide a brief overview of these demonstrations and discuss relevant findings from the analysis of subgroups conducted for SSDI beneficiaries and SSI recipients, respectively. (The Appendix in this volume provides additional information about all of SSA's demonstrations.)[7] The last subsection summarizes some broader lessons and practical insights from this discussion.

General Considerations for Comparing Results across Demonstrations

The eight randomized evaluations of SSA's demonstrations considered here all to some degree addressed potential differences in the impact of the evaluated intervention across groups of individuals. The dimensions of heterogeneity varied across studies. Exhibit 7.1 indicates the groups that each of the demonstrations analyzed. In terms of demographic differences, seven out of eight studies differentiated by age, four by education, four by gender, and two by race/ethnicity. Seven out of eight studies differentiated among types of health impairments, five among types of benefit receipt

---

[6] Following the Extended Period of Eligibility (or its reentitlement period), if SSDI payments have stopped because a beneficiary's income is substantial, SSA gives them five years during which their benefits can be reinstated if they again stop working because of their disability. During the five-year period, SSA will not require them to file a new disability application to get benefits; this is called Expedited Reinstatement. For those workers who lost their entitlement to benefits but need to quit working for the same or related medical impairment, Expedited Reinstatement allows benefits to start again without their needing to submit a new application. See, for example, https://choosework.ssa.gov/library/fact-sheet-trial-work-period-twp or the *Red Book* (SSA 2020e; https://www.ssa.gov/redbook).

[7] Additional analysis of the subgroups, subgroup impacts, and differential subgroup impacts discussed in this chapter is available on request from the author.

(i.e., SSDI-only and SSDI/SSI concurrent enrollments), and five between having prior employment or not. To aid the exposition, the chapter will refer to the higher-ordered group as "category" (e.g., gender, impairment type) and the defining characteristic within each category as "subgroup" (e.g., women, musculoskeletal system impairments).

Exhibit 7.1. Subgroups Included in the Analysis, by Demonstration

	Age	Gender	Education	Employment	Impairment	SSI or SSDI Receipt	Other
AB	X				X	X	X
BOND (Stage 1)	X			X	X	X	X
BOND (Stage 2)	X		X	X	X	X	X
BOPD	X	X	X	X			X
MHTS	X	X	X		X		
Project NetWork					X	X	
POD	X		X	X	X	X	X
STETS	X	X	X	X	X	X	X
TETD	X	X	X	X	X	X	X

Note: Depending on the demonstration, "Other" includes Medicaid use, health (self-reported), body mass index, race/ethnicity, location, living arrangements, future expectations.

When comparing differences in the effect of inducements and supports of employment among beneficiaries between studies, one potential difficulty is that each study has slightly different definitions of its main outcomes. The most consistently available outcome across the studies I review is the total amount of earnings and the total amount of SSDI benefits paid, corresponding to the focus of the demonstrations on testing policies aimed at reintegrating beneficiaries into the labor market. Other frequently examined outcomes were employment and incidence of earnings above the SGA level.

To achieve a minimum amount of comparability among studies, the discussion focuses on those two outcomes common across almost all studies: total earnings and total amount of SSDI benefits paid. Yet there are still some differences in how these outcomes are defined. One observation for SSA to consider: whether future demonstrations should have greater consistency of outcome measures used.

Exhibit 7.2 (beginning on page 2) summarizes subgroup impacts from among the SSDI-focused demonstrations, and Exhibit 7.3 (beginning on page 2) summarizes subgroup impacts from among the SSI-focused demonstrations. In each exhibit, Panel A reports earnings impacts and Panel B reports benefits impacts. It is important to note that only the BOND report provided standard errors for the difference in the estimated impacts within a category, information needed to report whether the difference in the impacts *between* two subgroups is statistically significant. In the remaining cases, we can only assess whether the finding *within* a particular subgroup is statistically significantly different from zero, but not whether there are statistically detectable

differences *between* subgroups' impacts. This can be an important drawback for understanding subgroup heterogeneity and is further discussed in the last subsection ("Potential Insights and Practical Considerations").

BOND, MHTS, STETS, and TETD provide extensive subgroup impact estimates; whereas AB and Project NetWork provide mainly a description of subgroup-related findings. BOPD had a more limited exploration of subgroups. As result, the discussion will focus on BOND, MHTS, STETS, and TETD, with shorter mention of the other demonstrations. Preliminary results from POD's Interim Evaluation Report are discussed briefly, as well.

Demonstrations Focused on SSDI and Concurrent Beneficiaries

Benefit Offset National Demonstration (BOND)[8]

Informed by the results of its pilot study (BOPD), BOND tested the impact of a benefit offset on a nationally representative sample of SSDI (or concurrent SSDI/SSI) beneficiaries. In a first stage, the evaluation randomly assigned all SSDI beneficiaries in 10 randomly chosen SSA areas to either a treatment group that receives the offset and work incentives counseling (WIC) or a control group that receives only WIC. A second stage tested the impact of the offset on a group of SSDI-only beneficiaries who volunteered for the demonstration and thus were expected to be more likely to use the offset. In addition, a second stage tested the extent to which enhanced work counseling improves outcomes compared to WIC services by randomly assigning volunteers to either treatment 1 (benefit offset and WIC), treatment 2 (benefit offset and enhanced WIC), or a control group. All treated participants had access to the benefit offset during a 60-month participation period after completing the TWP.

---

[8]   Discussion based on Gubits et al. (2018a/b).

Exhibit 7.2. Summary of Subgroup Analysis of Demonstrations Focused on SSDI Recipients

			Panel A: Earnings			
Age	Education	Employment Status	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	SSDI Benefit Duration	Gender
BOND (Stage 1): Total Earnings	N/A					N/A
Age ≤49		Employed	Affective disorder	SSDI	Short	
T1: $9,328		T1: $28,954	T1: $7,336	T1: $7,276	T1: $7,687	
C1: $9,336		C1: $29,295	C1: $7,127	C1: $7,245	C1: $7,754	
Impact: –$8		Impact: –$341	Impact: $209	Impact: $31	Impact: –$67	
Age >49		Not employed	Other impairment[a]	Concurrent	Long	
T1: $4,212		T1: $2,026	T1: $6,499	T1: $3,724	T1: $6,175	
C1: $4,188		C1: $1,946	C1: $6,528	C1: $3,812	C1: $6,133	
Impact: $24		Impact: $80	Impact: –$29	Impact: –$88	Impact: $42	
Diff: –$32		Diff: –$421	Diff: $238	Diff: $119	Diff: –$109	
			Back disorder			
			T1: $4,819			
			C1: $4,724			
			Impact: $95			
			Other impairment[b]			
			T1: $6,929			
			C1: $6,933			
			Impact: –$5			
			Diff: $100			

BOND (Stage 2): Total Earnings

Age	Education	Employment Status	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	SSDI Benefit Duration	Gender
Age ≤49	<Associate's	Employed	Affective disorder	SSDI	Short	N/A
T21 & T22: $22,661	T21 & T22: $15,746	T21 & T22: $42,374	T22 + T21: $20,037	T1: $19,522	T21 + T22: $18,968	
C2: $20,690	C2: $13,797	C2: $39,870	C1: $18,079	C1: $17,348	C2: $17,348	
Impact: $1,971*	Impact: $1,948*	Impact: $2,504	Impact: $1,958	Impact: $2,174	Impact: $1,620*	
Age >49	Any postsecondary degree	Not employed	Other impairment[a]	Concurrent	Long	
T21 & T22: $13,464	T21 & T22: $22,724	T21 & T22: $9,618	T22 + T21: $17,573	T1: $17,109	T21 + T22: $17,383	
C2: $12,355	C2: $21,665	C2: $8,236	C1: $16,136	C1: $15,901	C2: $15,901	
Impact: $1,109	Impact: $1,059	Impact: $1,382	Impact: $1,437*	Impact: $1,207	Impact: $1,482	
Diff: $862	Diff: $889	Diff: $1,123[c]	Diff: $521	Diff: $966	Diff: $138	
			Back disorder			
			T22 + T21: $13,378			
			C2: $13,849			
			Impact: –$471			
			Other impairment[b]			
			T22 + T21: $18,743			
			C2: $16,904			
			Impact: $1,839*			
			Diff: –$2,310			

Age	Education	Employment Status	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	SSDI Benefit Duration	Gender
MHTS: Past 3 Month's Earnings (at study exit)						
Age 18-34 T: $913 C: $879 Impact: n.r. Age >35 T: $854 C: $443 Impact: n.r.* Diff: n.r.	<High school T: $747 C: $378 Impact: n.r.* High school T: $754 C: $328 Impact: n.r.* >High school T: $923 C: $568 Impact: n.r.* Diff: n.r.	N/A	Affective disorder T: $958 C: $500 Impact: n.r.* Schizophrenia T: $639 C: $421 Impact: n.r.* Diff: n.r.	SSDI only T: $2,374 C: $2,048 Impact: $326 SSI only T: $1,060 C: $893 Impact: $167 Concurrent T: $1,249 C: $1,279 Impact: –$30 Diff: n.r.	N/A	Men T: $862 C: $402 Impact: n.r.* Women T: $855 C: $550 Impact: n.r.* Diff: n.r.
POD: Earnings in 2019						
Age <50 T: $6,059 C: $6,131 Impact: –$71 Age ≥50 T: $3,711 C: $3,728 Impact: –$17 Diff: n.s.	≤High school T: $4,187 C: $4,209 Impact: -$22 >High school T: $5,929 C: $6,048 Impact: –$119 Diff: n.s.	Employed T: $13,416 C: $13,914 Impact: –$499 Not employed T: $2,217 C: $2,063 Impact: $154 Diff: n.s.	Mental T: $4,817 C: $4,752 Impact: $64 Musculoskeletal T: $4,398 C: $4,366 Impact: $32 Other[d] T: $5,138 C: $5,340 Impact: –$202 Diff: n.s.	N/A	N/A	N/A

Panel B: SSDI Benefits

Age	Education	Employment Status	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	SSDI Benefit Duration
BOND (Stage 1): Total Benefits Due					
Age ≤49	N/A	Employed	Affective disorder	SSDI	Short
T1: $46,794		T1: $57,704	T1: $53,748	T1: $60,066	T1: $58,292
C1: $45,845		C1: $55,146	C1: $53,207	C1: $59,422	C1: $57,829
Impact: $949*		Impact: $2,558*	Impact: $541*	Impact: $644*	Impact: $463*
Age >49		Not employed	Other impairment[a]	Concurrent	Long
T1: $60,758		T1: $53,423	T1: $54,231	T1: $27,511	T1: $52,368
C1: $60,398		C1: $53,162	C1: $53,576	C1: $26,903	C1: $51,655
Impact: $359*		Impact: $261*	Impact: $655*	Impact: $608*	Impact: $714*
Diff: $590†		Diff: $2,297†	Diff: –$114	Diff: $37	Diff: –$251
			Back disorder		
			T1: $64,252		
			C1: $63,790		
			Impact: $462*		
			Other impairment[b]		
			T1: $52,496		
			C1: $51,829		
			Impact: $667*		
			Diff: –$205		

BOND (Stage 2): Total Benefits Due

Age	Education	Employment Status	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	SSDI Benefit Duration
Age ≤49	<Associate's	Employed	Affective disorder	SSDI	Short
T21 & T22: $46,543	T21&T22: $48,664	T21 & T22: $47,863	T22 + T21: $49,292	T1: $53,965	T21 & T22: $53,958
C2: $44,373	C2: $47,049	C2: $43,973	C2: $48,210	C1: $52,361	C2: $52,361
Impact: $2,170*	Impact: $1,615*	Impact: $3,890*	Impact: $1,083	Impact: $1,604*	Impact: $1,597*
Age >49	Any postsecondary degree	Not employed	Other impairment[a]	Concurrent	Long
T21 & T22: $56,434	T21 & T22: $57,421	T21 & T22: $52,825	T22 + T21: $52,033	T1: $49,928	T21 & T22: $49,724
C2: $54,861	C2: $55,116	C2: $51,644	C2: $49,971	C1: $47,652	C2: $47,652
Impact: $1,574*	Impact: $2,305*	Impact: $1,180*	Impact: $2,062*	Impact: $2,277*	Impact: $2,072*
Diff: $597	Diff: –$690	Diff: $2,709†	Diff: –$979	Diff: –$672	Diff: –$476
			Back disorder		
			T22 + T21: $54,175		
			C2: $53,329		
			Impact: $846		
			Other impairment[b]		
			T22 + T21: $51,096		
			C2: $49,085		
			Impact: $2,011*		
			Diff: –$1,165		

Age	Education	Employment Status	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	SSDI Benefit Duration
MHTS: Past Month's SSDI Benefit Amount					
Age 18-34	<High school	N/A	Affective disorder	SSDI only	N/A
T: $689	T: $748		T: $896	T: $630	
C: $680	C: $701		C: $887	C: $628	
Impact: n.r.	Impact: n.r.		Impact: n.r.	Impact: $2	
Age ≥35	High school		Schizophrenia	SSI only	
T: $871	T: $800		T: $767	T: $36	
C: $868	C: $790		C: $759	C: $42	
Impact: n.r.	Impact: n.r.		Impact: n.r.	Impact: –$6	
Diff. n.r.	>High school		Diff. n.r.	Concurrent	
	T: $899			T: $378	
	C: $913			C: $383	
	Impact: n.r.			Impact: –$5	
	Diff. n.r.			Diff. n.r.	

POD: SSDI Benefit Amount (in 12 months after enrollment)

Age	Education	Employment Status	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	SSDI Benefit Duration
Age <50	>High school	Employed	Mental	N/A	N/A
T: $11,037	T: $13,248	T: $10,641	T: $11,229		
C: $10,997	C: $13,447	C: $10,850	C: $11,125		
Impact: $40	Impact: -$199	Impact: -$209	Impact: $103		
Age ≥50	≤High school	Not employed	Musculoskeletal		
T: $12,890	T: $11,214	T: $12,401	T: $12,738		
C: $12,956	C: $11,092	C: $12,352	C: $12,656		
Impact: -$66	Impact: $122	Impact: $49	Impact: $82		
Diff: n.s.	Diff: n.s.	Diff: n.s.	Other impairment[d]		
			T: $12,338		
			C: $12,489		
			Impact: -$152		
			Diff: n.s.		

Notes: Due to rounding, values that were computed at greater levels of precision may appear not to sum as whole numbers.

The "Impact" is the treatment-control difference; the "Diff" is the between-group difference in subgroup impacts.

n.r. indicates that the evaluation did not report the value (but in some cases, it did report whether the value was statistically significantly different from zero and is noted).

n.s. indicates that the diff value, not reported, is not statistically significantly different from zero.

Statistical significance is indicated (per the specific evaluation's conventions) as follows: * = impact is statistically significantly different from zero. † = diff (impact differential between subgroups) is statistically significantly different from zero.

[a] "Other impairments" denotes everything other than affective disorder.

[b] "Other impairments" denotes everything other than back disorder.

[c] *p*-value was 10.3%.

[d] "Other impairments" denotes everything other than mental disorder or musculoskeletal disorder.

BOND was notable due to its large sample sizes. The primary outcomes of the study were a cumulative earnings measure (2011–2015 for Stage 1 and 2012 and 2015 for Stage 2) and total SSDI benefits (as recorded in May 2017).[9] BOND did not find any evidence of the benefit offset policy increasing total earnings or decreasing total SSDI benefits in either stage. In fact, the evaluation found strong evidence of the offset policy increasing SSDI benefits in both stages.[10] The different forms of work incentives counseling in Stage 2 treatments did not have any effect on earnings or SSDI benefits.

BOND is by far the largest experimental demonstration reviewed here, and therefore its analysis would be expected to be most likely to detect differences across groups. As noted earlier in Exhibit 7.1, Stage 1 of the demonstration evaluated the effect of treatment across the following subgroups: age, employment status, type of health impairment, SSDI benefit duration, SSI status, and access to a Medicaid buy-in program. Stage 2 of the demonstration evaluated impacts for a slightly different set of subgroups: age, employment status, type of health impairment, SSDI benefit duration, and access to a Medicaid buy-in program.

Looking across estimates in Exhibit 7.2 of the effect of the benefit offset on earnings from the BOND study, it appears that overall there are no cases in which impact estimates are found to be statistically different within categories. The only *difference* in subgroup effects within categories that approaches significance relates to impact estimates by prior employment in Stage 2. Those who were not employed at baseline had an earnings impact, whereas those who were employed at baseline did not. The difference between these impacts approaches statistical significance (with a *p*-value of .103). Next, in Stage 2 of BOND, it appears that impact estimates for younger beneficiaries (age 49 and younger), beneficiaries with less than an associate's degree, beneficiaries in the subgroup with a primary impairment other than a major affective disorder, and beneficiaries in the subgroup with a primary impairment other than a back disorder were found to be positive and statistically significantly different from zero. (No subgroup estimates were statistically significantly different from zero in Stage 1 of BOND.)

Although not statistically significant, the differences in impact estimates between some of the other subgroups are substantial—for example, younger beneficiaries have

---

9   Previous BOND reports used a different outcome measure, SSDI benefits paid. SSA occasionally makes incorrect payments to beneficiaries and later corrects for these payments. "Benefits paid" is the value SSA paid a beneficiary at the time; "benefits due" is a revised measure of the amount a beneficiary should have received at the time.

10 This can be explained by the combination of a larger positive mechanical effect and a smaller negative behavioral effect on SSDI benefits. The second of these effects (which implies individuals moving from full benefits to partial benefits because of increased employment) is swamped by a larger amount of individuals already working at the SGA level who mechanically go from zero benefits (due to suspense under current-law rules) to partial benefits (because of the benefit offset). Given the structure of the offset, there needed to be more beneficiaries moving into SGA from nonemployment for average benefits to go down.

nearly double the earnings increase of older beneficiaries, and the same is true for more- versus less-educated beneficiaries. These are important findings, as they show that in contrast to the zero average effect, certain salient subgroups appear to have experienced increases in earnings in response to the benefit offset.

Considering effects on receipt of SSDI benefits in Panel B of Exhibit 7.2, almost all subgroup-specific estimates are statistically significant in both stages of BOND. These effects are all positive, indicating that the benefit offset raised rather than lowered SSDI benefit amounts.[11] In three instances, the differences within categories are statistically significant. For example, in both stages, we can see that beneficiaries with prior employment (compared to those without) have substantially larger increases in SSDI benefits (both in absolute and percentage terms). Similarly, there is a difference between impact estimates for older and younger people in Stage 1, but not Stage 2.

Interestingly, for both the employment status and the age group comparisons, the groups with higher earnings impacts also have higher impacts on SSDI benefits received. In contrast, for education groups or SSDI-only versus concurrent beneficiaries, those groups with higher earnings impacts have lower (albeit still positive) impacts on SSDI benefits.

The BOND study also evaluated subgroup impacts for other outcomes. In total, 364 tests of difference in impacts were conducted for Stage 2 subgroup analysis, implying that some of the tests would be statistically significant by chance. Yet, there was no clear pattern of the offset's behavioral effects in the subgroup analysis beyond those already discussed. Weak evidence is presented that the effect on employment and earnings above the SGA level outcomes is greater for participants with less education (statistically significant in 2 out of 12 tests; see Gubits et al. [2018b, Exh. F-49, F-50, and F-52]). This subgroup had lower rates of employment overall, which the report suggests could have led to larger effects for this group.

Mental Health Treatment Study (MHTS)[12]

The MHTS tested how access to supported employment services and systematic medication management services affects the ability of SSDI beneficiaries with schizophrenia or an affective disorder to return to work. The treatment group received a comprehensive package of mental health and employment services and was exempted from medical continuing disability reviews for a three-year period after study enrollment. The control group was given a list of available local and national resources along with a $100 payment for participating in quarterly interviews and was not exempted from medical continuing disability reviews. Relative to BOND or Project NetWork, MHTS was a relatively small evaluation, involving 2,238 volunteers

---

[11] This is consistent with the intent of BOND to leave participants on average better off than nonparticipants by allowing them to keep receiving some benefits while working.

[12] Discussion based on Frey et al. (2011).

among SSDI beneficiaries between the ages of 18 and 55 diagnosed with either schizophrenia or an affective disorder from 23 study sites. Once enrolled, participants remained in the study for two years.

Given MHTS's focus on employment and health outcomes for SSDI beneficiaries, the primary outcomes of interest were a participant's monthly employment rate, self-reported physical and mental health scores, and quality of life. The study had a number of other exploratory outcomes related to employment and health. MHTS's intervention had substantial positive effects on the employment rate and earnings but did not lead to a statistically significant reduction in SSDI benefits. The study found that the mental health score (but not the physical health score) and general life satisfaction improved for the treatment group relative to the control group during the study period.

MHTS is another demonstration that reports detailed subgroup impacts. As shown earlier in Exhibit 7.1, the demonstration evaluated treatment impacts for the following subgroups: age, gender, educational attainment, and disorder diagnosis. The report did not provide test statistics that would allow us to assess whether impact estimates across subgroups within categories were statistically different from one another. Comparing between earnings (Panel A) and SSDI receipt (Panel B), for MHTS the opposite pattern from BOND emerges. There are no detectable subgroup impacts for SSDI receipt; however, there are several instances of subgroup impacts for earnings.

MHTS distinguishes between earnings averages that consider the whole sample (i.e., include zeros for those who are not employed; called "unconditional" estimates) and averages that consider those who are employed (i.e., exclude zeros for those who are not employed; called "conditional" estimates). Because the observed impacts on employment imply the estimates of the conditional earnings could be based on a selected sample of workers, this chapter reports just the unconditional (experimentally valid) impacts.

These earnings impacts appear to be larger for older workers than for younger workers, the latter being the only group that does not have a detectable increase. All three of the education groups considered experienced earnings impacts. These earnings impacts are positive (and precisely estimated) also for those with affective disorder and schizophrenia, with the former group experiencing somewhat larger increases.

Overall, these impact estimates suggest the treatment was broadly successful in raising employment for the population of eligible beneficiaries, with young beneficiaries being a clear exception. As confirmed by Panel B, in none of the subgroups did the rise in earnings lead to a reduction in SSDI benefits.

Other SSDI-Related Demonstrations

In contrast to BOND and MHTS, BOPD, and Project NetWork did not engage in a systematic analysis of subgroup impacts. The remainder of the section summarizes some of the results from these three evaluations as discussed in the respective reports. Preliminary results from POD are mentioned, as well.

Benefit Offset Pilot Demonstration (BOPD).[13] The four-state BOPD was the pilot study for BOND. BOND tested the effect of an intervention that reduced annual SSDI benefits by $1 for every $2 of annual earnings above an annualized measure of SGA. By providing beneficiaries with a "ramp" that gradually reduces SSDI benefits as earnings increase, the benefit offset treatment prevented SSDI beneficiaries (or concurrent SSDI/SSI beneficiaries) earning more than the SGA amount from facing a "cash cliff" after the TWP runs out. From 2004 to 2010, BOPD randomly assigned a group of volunteers from four states (CT, UT, VT, WI) to either a treatment group (N=917) or a control group (N=893). The benefit offset was available to volunteers assigned to the treatment group for a six-year period after they completed the nine-month TWP. The pilot focused on working through issues in administering the offset, but it also measured the impact of the offset on average annual earnings, the incidence of having any employment in a given year, and annual SSDI benefits received by volunteers as measured by administrative records.

The benefit offset increased the proportion of beneficiaries with earnings above the SGA level in the two years after random assignment (Weathers and Hemmeter 2011). However, the benefit offset did not result in reductions in SSDI benefit payments. Moreover, among beneficiaries who made more than the SGA level before random assignment, the benefit offset provisions reduced average earnings.

Each state also conducted its own evaluation, including subgroup analyses. For example, in Wisconsin, subgroup analysis was conducted for six pairs, including analysis of age, gender, earnings history (any earnings and $1,200 cutoff), Medicaid buy-in,[14] and TWP completion (Delin et al. 2010). In terms of annual earnings, out of 108 subgroups, the 5 for which there was an impact were women, the "no Medicaid buy-in" group, those with "pre-enrollment earnings," and those with "less than $1,200 pre-enrollment earnings." All of these impacts occurred in the first and second quarter after study enrollment, with no detectable impacts from the third quarter to the eighth quarter.

Connecticut and Vermont exhibited similar trends of early impacts for the Medicaid buy-in subgroup, and Connecticut also reported impacts in the older subgroup (Porter et al. 2009; State of Connecticut 2009). Although there were no detectable impacts for the Medicaid buy-in or age subgroups in Utah, there were impacts for men and subgroups based on pre-program earnings (Chambless et al. 2009).

The subgroup analysis did not examine impacts on SSDI benefits. With the exception of the Medicaid buy-in distinction, the subgroup analyses do not change the impression from BOND, which was a larger experiment that tested similar subgroups. BOPD's subgroup analyses do reflect, however, the presence of differences across treatment sites also seen in other demonstrations.

---

[13] Discussion based on Weathers and Hemmeter (2011).

[14] Medicaid buy-in programs allow workers with disabilities access to Medicaid services.

Project NetWork.[15] The Project NetWork demonstration studied the impact that case management had on earnings and SSDI receipt for beneficiaries with severe disabilities. Implemented in 1991 at eight sites across the United States, Project NetWork recruited 8,248 volunteers of eligible SSDI (or concurrent SSDI/SSI) beneficiaries. Participants were randomly assigned into the treatment group or the control group. The demonstration provided treatment group members with three services: outreach, waivers, and case management. The study used four different case management models. All models had the same outreach procedures and work incentive waiver provisions but differed in the implementation of case management. Each model was implemented in two of the eight sites.

Project NetWork measured the impact of treatment on average monthly disability benefits, receipt of services, and average earnings two years after study enrollment. The program overall achieved annual earnings gains in the first two follow-up years (but not in the third). Project NetWork did not reduce the amount of SSDI or SSI receipt.

The Project NetWork evaluation analyzed impacts for subsets of the sample defined by type of eligibility (SSDI-only versus SSI-only versus concurrent) and primary impairment (mental versus neurological versus musculoskeletal versus other). Average annual earnings increased for SSDI-only beneficiaries. SSI-only recipients, and concurrent beneficiaries did not have a detectable earnings impact, which could reflect prior work experience or imply that SSDI-only beneficiaries required fewer services in order to return to work.

Important to this chapter is that no test for the difference between groups' impact estimates was provided. Discussion of differences in impact estimates by impairment type noted that treatment generally did not affect earnings or SSDI benefits for any of the impairment types considered (Kornfeld and Rupp 2000).

Accelerated Benefits (AB).[16] The AB demonstration tests how early access to medical services affects new SSDI beneficiaries' health and employment outcomes. The demonstration provided new SSDI beneficiaries with access to AB health care during the 24-month waiting period before their transition to Medicare. The demonstration randomly assigned 1,997 participants into three groups: in addition to standard SSDI benefits, treatment group 1 (AB) received access to AB health care whereas treatment group 2 (AB Plus) received AB health care plus telephone counseling services; the control group received only SSDI benefits.

The primary outcomes focused on three health factors: health care use, unmet need, and health status. The exploratory variables focused on employment. The main results for the primary outcomes of the demonstration were a rise in health care utilization and a reduction in the share of participants with unmet medical needs. Although the AB Plus group tended to look more for work and used more Ticket to

---

[15] Discussion based on Kornfeld and Rupp (2000).

[16] Discussion based on Michalopoulos et al. (2011).

Work and vocational services compared to the AB or control group, the intervention was not shown to have any effect on employment outcomes.

Subgroup analysis was conducted by type of impairment and age. For participants with mental health impairments, both treatment groups experienced statistically insignificant decreases in the share ever employed, compared to the control group. However, participants with "other impairments" increased employment, with the AB Plus group having an impact of +3.4 percentage points (the AB group had a +3.8 percentage point change, though not statistically significant).

Promoting Opportunity Demonstration (POD). Building off BOND and BOPD, POD tests new benefit offset rules that simplify work incentives to promote employment and reduce administrative complexity. The simplified POD rules eliminated the TWP and Grace Period, and they used a uniform benefit offset formula. The demonstration varied rules about SSDI benefit termination across two treatment arms. The first treatment group (T1) could not have their benefits terminated for work, whereas the second treatment group (T2) could have their benefits terminated if participants were in full offset for 12 consecutive months. Implemented in eight states, POD enrolled 10,070 participants, randomly assigning 3,343 participants to T1 and 3,357 participants to T2. The simplified POD rules appear to have driven higher use of the benefit offset, with 24 percent of POD treatment participants (both T1 and T2) using the offset one year after enrollment, compared to only 7 percent of BOND treatment participants.

Impact estimates discussed here come from POD's interim report and compare the control group to pooled treatment groups. As of one year after enrollment, POD had no detectable impact on any of the primary outcomes. With one-quarter of treatment group members using the benefit offset, POD rules should mechanically increase benefit payments among some treatment group members. The lack of detectable impacts suggests that increases in benefit payments were offset by decreases. POD also examined various subgroup impacts, including those by age, education, employment status, and impairment type. There were no statistically significant differences within or across groups.

Demonstrations Focused on SSI Recipients

Two demonstrations that focused on SSI recipients—STETS and TETD—examined program impacts by subgroups. The subgroup-specific impact estimates for these two demonstrations are summarized in Exhibit 7.3 below.

Structured Training and Employment Transitional Services (STETS)[17]

Testing the impact of transitional employment services, the STETS demonstration offered training and transitional job placement services to young people (ages 18–24)

---

[17] Discussion based on Kerachsky et al. (1985).

with intellectual disability (IQ scores between 40 and 80) and limited prior work experience.[18] Funded by the US Department of Labor, the STETS demonstration implemented a transitional employment model that consisted of three phases: Phase 1 provided training and support in a low-stress work environment; Phase 2 transitioned participants to on-the-job training at local businesses in a regular work environment; Phase 3 consisted of follow-up services to those who had transitioned to competitive jobs. Program participation in Phases 1 and 2 was expected to last for roughly 12 months.

STETS operated from fall 1981 through December 1983 and implemented the transitional employment model in five locations—Cincinnati, OH; Los Angeles, CA; New York, NY; St. Paul, MN; and Tucson, AZ. The STETS demonstration enrolled 437 participants and randomly assigned 226 participants to the treatment group and the remaining 211 participants to the control group. The primary outcomes for the demonstration were employment, income, SSI receipt, and service use.

For the full sample, the evaluation of STETS found no impact of the treatment (transitional employment services) on weekly personal income of young SSI recipients with intellectual disability. It did find an increase in the fraction of recipients working in regular competitive jobs in the labor market (as opposed to working in any job, which includes training jobs that were part of the treatment). There was no detectable effect on SSI benefits received.

Subgroup analyses were conducted for the two primary outcomes—weekly income and average monthly SSDI/SSI benefits—for a range of subgroups shown in Exhibit 7.3. With respect to earnings, the subgroup analysis indicates that the treatment increased weekly income among individuals with moderate intellectual disability by $43.40, compared to the control group. It had no effect for those with borderline or mild intellectual disability. With respect to receipt of benefits, the treatment reduced receipt of SSDI/SSI among those who received other transfers (including any cash transfers and Medicaid) or no transfers at baseline by $60 monthly, compared to the control group.

Impacts on the employment outcome—percentage employed in a regular job—occurred for subgroups defined by race/ethnicity. Hispanic treatment group members experienced a 22.6 percentage point increase in their employment in a regular job, and the White non-Hispanic/Other subgroup experienced a 10.5 percentage point increase. There was not a detectable impact for Black non-Hispanic group members (their roughly 10 percentage point increase was not statistically different from zero).

---

[18]   Though the demonstration's target population exhibited high dependence on others (measured through living arrangements), only one-third of participants were receiving either SSI or SSDI benefits.

Exhibit 7.3. Summary of Subgroup Analysis of Demonstrations Focused on SSI Recipients

			Panel A: Earnings			
Age	Education	Employment Status (prior work experience)	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	Gender	Race/Ethnicity
STETS: Average Weekly Personal Income						
Age <22	Enrolled	Regular job lasting 3+ months	Borderline	SSI/SSDI transfer	Men	Black non-Hispanic
T: $69	T: $67	T: $85	T: $76	T: $98	T: $80	T: $63
C: $61	C: $59	C: $84	C: $58	C: $73	C: $63	C: $56
Impact: $8	Impact: $8	Impact: <$1	Impact: $18	Impact: $25*	Impact: $17*	Impact: $8
Age ≥22	Not enrolled	Any job lasting 3+ months	Mild	Other transfer	Women	Hispanic
T: $78	T: $74	T: $72	T: $66	T: $70	T: $60	T: $87
C: $66	C: $64	C: $62	C: $67	C: $60	C: $61	C: $65
Impact: $12	Impact: $10	Impact: $10	Impact: −$1	Impact: $10	Impact: −$1	Impact: $22
Diff: n.r.	Diff: n.r.	Other	Moderate	No transfer	Diff: n.r.	White non-Hispanic and Other
		T: $68	T: $90	T: $46		T: $72
		C: $56	C: $47	C: $54		C: $65
		Impact: $11	Impact: $43*	Impact: −$8		Impact: $7
		Diff: n.r.	Diff: n.r.	Diff: n.r.		Diff: n.r.

TETD: Average Earnings after First 24 Months

Age	Education	Employment Status (prior work experience)	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	Gender	Race/Ethnicity
Age <22	N/A	Regular job	IQ score <40	Received SSI	Men	Black non-Hispanic
T: $2,847		T: $4,153	T: $1,939	T: $3,315	T: $3,092	T: $3,173
C: $1,635		C: $3,186	C: $710	C: $1,389	C: $1,430	C: $1,180
Impact: $1,212*		Impact: $967	Impact: $1,229	Impact: $1,926*	Impact: $1,662*	Impact: $1,993*
Age ≥22		Training/volunteer job	IQ score 40-54	Did not receive	Women	White non-Hispanic and Other
T: $3,217		T: $2,934	T: $2,961	T: $3,052	T: $3,196	T: $3,102
C: $1,533		C: $904	C: $1,655	C: $1,627	C: $1,738	C: $1,699
Impact: $1,684*		Impact: $2,030*	Impact: $1,306*	Impact: $1,425*	Impact: $1,458*	Impact: $1,403*
Diff. n.r.		Sheltered workshop	IQ score 55-70	Diff. n.r.	Diff. n.r.	Diff. n.r.
		T: $3,758	T: $3,125			
		C: $1,989	C: $1,557			
		Impact: $1,769*	Impact: $1,568*			
		Other job	IQ score >70			
		T: $2,984	T: $4,108			
		C: $1,109	C: $1,296			
		Impact: $1,874*	Impact: $2,811*			
		No job	Diff. n.r.			
		T: $2,206				
		C: $905				
		Impact: $1,301*				
		Diff. n.r.				

Panel B: SSI Benefits

STETS: Average Monthly Income from SSI or SSDI

Age	Education	Employment Status (prior work experience)	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	Gender	Race/Ethnicity
Age <22	Enrolled	Regular job lasting 3+ months	Borderline	SSI/SSDI transfer	Men	Black
T: $86	T: $117	T: $96	T: $90	T: $203	T: $91	T: $67
C: $112	C: $132	C: $141	C: $83	C: $181	C: $134	C: $93
Impact: –$26	Impact: –$15	Impact: –$45	Impact: $6	Impact: $23	Impact: –$42*	Impact: –$25
Age ≥22	Not enrolled	Other job lasting 3+ months	Mild	Other transfer	Women	Hispanic
T: $129	T: $91	T: $107	T: $98	T: $54	T: $109	T: $94
C: $138	C: $115	C: $88	C: $144	C: $114	C: $102	C: $95
Impact: –$9	Impact: –$23	Impact: $19	Impact: –$46*	Impact: –$60*	Impact: $7	Impact: <(–$1)
Diff. n.r.	Diff. n.r.	Other	Moderate	No transfer	Diff. n.r.	White and other
		T: $95	T: $132	T: $36		T: $118
		C: $136	C: $85	C: $64		C: $142
		Impact: –$40*	Impact: $47	Impact: –$28		Impact: –$24
		Diff. n.r.	Diff. n.r.	Diff. n.r.		Diff. n.r.

Age	Education	Employment Status (prior work experience)	Type of Health Impairment	SSDI vs. SSI vs. Concurrent	Gender	Race/Ethnicity
STETS: Average Monthly Income from SSI or SSDI						
<u>Age <22</u> T: $86 C: $112 Impact: –$26 <u>Age ≥22</u> T: $129 C: $138 Impact: –$9 Diff: n.r.	<u>Enrolled</u> T: $117 C: $132 Impact: –$15 <u>Not enrolled</u> T: $91 C: $115 Impact: –$23 Diff: n.r.	<u>Regular job lasting 3+ months</u> T: $96 C: $141 Impact: –$45 <u>Other job lasting 3+ months</u> T: $107 C: $88 Impact: $19 <u>Other</u> T: $95 C: $136 Impact: –$40* Diff: n.r.	<u>Borderline</u> T: $90 C: $83 Impact: $6 <u>Mild</u> T: $98 C: $144 Impact: –$46* <u>Moderate</u> T: $132 C: $85 Impact: $47 Diff: n.r.	<u>SSI/SSDI transfer</u> T: $203 C: $181 Impact: $23 <u>Other transfer</u> T: $54 C: $114 Impact: –$60* <u>No transfer</u> T: $36 C: $64 Impact: –$28 Diff: n.r.	<u>Men</u> T: $91 C: $134 Impact: –$42* <u>Women</u> T: $109 C: $102 Impact: $7 Diff: n.r.	<u>Black non-Hispanic</u> T: $67 C: $93 Impact: –$25 <u>Hispanic</u> T: $94 C: $95 Impact: <(–$1) <u>White non-Hispanic and Other</u> T: $118 C: $142 Impact: –$24 Diff: n.r.

Notes: Due to rounding, values that were computed at greater levels of precision may appear not to sum as whole numbers.

The "Impact" is the treatment-control difference; the "Diff" is the between-group difference in subgroup impacts.

n.r. indicates that the evaluation did not report the value (but in some cases, it did report whether the value was statistically significantly different from zero and reported and is noted).

n.s. indicates that the diff value, not reported, is not statistically significantly different from zero.

Statistical significance is indicated (per the specific evaluation's conventions) as follows:

* Impact is statistically significantly different from zero

† Diff (impact differential between subgroups) is statistically significantly different from zero.

STETS also saw substantial differences across treatment sites. For example, the treatment group in St. Paul (MN) experienced a 23 percentage point increase in employment in a regular job. The treatment group in Cincinnati (OH) saw a nearly 15 percentage point increase in employment in any paid job. On earnings from regular jobs, the intervention increased earnings in all sites, but in only St. Paul and Los Angeles (CA) were those impacts statistically significant, at around $25 per week.

Transitional Employment Training Demonstration (TETD)[19]

The goal of TETD was to assess whether transitional employment services are effective at supporting SSI recipients with intellectual disability to gain economic self-sufficiency and maintain employment in competitive jobs. To test the impact of transitional employment services, TETD solicited volunteers from SSI recipients with intellectual disability ages 18–40 in 13 areas where the demonstration was taking place. Of 745 applicants, TETD randomly assigned participants into the treatment group (N=375) or the control group (N=370). The treatment group was offered access to transitional employment services, which consisted of three core services: placement in a "competitive" job; specialized on-the-job training; and post-placement support for job retention.

In a sign that the treatment was effective, the demonstration was able to place two-thirds of the treatment group in jobs, with half of them (or one-third of the treatment group) maintaining employment in the jobs, a rate consistent with other transitional employment programs. These services were time limited and were available to participants for only one year after enrollment in the study. The control group could not access these services but could access other services generally provided to SSI recipients.

The primary outcomes for the demonstration were employment, income, SSI receipt, and service use. The total earnings of the treatment group nearly doubled compared to the control group's, with the treatment group averaging slightly more than $3,100 and the control group averaging less than $1,600 during the first 24 months after enrollment. The increase in earnings is due to an increase in employment, with the share of the treatment group receiving any earnings being 18 percentage points greater than that of the control group. Though the demonstration increased earnings, it led to only small reductions in the average SSI benefits payment to treatment group members, with total SSI payments dropping by 4 percent ($266) over the 24-month period.

TETD engaged in extensive analysis of potential differences in treatment outcomes by subgroup. The subgroups were based on demographic (age, race/ethnicity, gender) and personal characteristics (IQ, motivation, physical ability); prior experiences (living arrangements, work experience, Social Security benefit receipt); and program services received. STETS had found that treatment was effective

---

[19]   Discussion based on Thornton and Decker (1989).

at raising income for participants with more severe intellectual disability, defined as IQ scores between 36 and 51. By contrast, TETD found that participants with less severe intellectual disability had higher earnings in the treatment group than in the control group. Only participants with an IQ score below 40 did not experience increases in earnings, with the treatment raising earnings for all the other participants.

Prior work experience before TETD enrollment also produced different impacts across groups. The impact on earnings for participants with less work experience was large, whereas that was not the case for participants with prior work experience in a regular job. This finding indicates the demonstration might have effectively transitioned participants without formal work experience into the mainstream workforce and increased their earnings.

Finally, although both young and old participants in the treatment group saw increased earnings, those older than age 22 experienced a larger increase in earnings than younger participants did. TETD did not provide a test of whether the subgroups' impacts were statistically significantly different from each other.

Like STETS, TETD exhibited some differences in the estimated treatment effects by site. Some sites implementing its intervention produced better outcomes than others did. Three project sites were particularly successful at increasing average earnings of treatment group members. One of them more than doubled the average earnings of the treatment group (an increase of about $2,000 in annual earnings) over the three-year period studied. That program placed participants in light manufacturing and assembly jobs.

Potential Insights and Practical Considerations

Potential Insights from Subgroup Impact Estimates

The demonstrations reviewed here vary in the outcomes studied and subgroups considered, partly due to the interventions and populations studied, partly due to practical considerations further discussed below. Nevertheless, some overarching insights emerge.

Overall, it is apparent that in many instances, the subgroup analysis reaffirms the main impact estimates. When impacts of the intervention are present, these are often reflected across subgroups (e.g., BOND for SSDI benefits; MHTS and TETD for earnings). Similarly, when no impact of the intervention is found for the overall population studied, most subgroup estimates are not statistically significantly different from zero, as well (e.g., BOND, BOPD, POD, and STETS for earnings; MHTS for SSDI benefits).

Yet, in some important cases, subgroup estimates diverge from the main impact estimates. For example, in BOND, the main impact estimate on earnings is not significantly different from zero. However, in Stage 2 of BOND, participants with "other impairments" (other than major affective disorder or a musculoskeletal disorder) and younger beneficiaries experienced increases in earnings. Similar patterns

occur for other demonstrations for which the main impact estimate is not statistically significantly different from zero (e.g., BOPD, STETS). As further discussed in the next main section, Recent Advances, such findings can be useful for deciding whether additional research is warranted to further explore variants of the intervention for the relevant subgroups.

In contrast, in MHTS, most subgroups experienced the overall increase in earnings, with the exception of younger workers, who did not experience an increase. Similar exceptions are observed for other demonstrations for which the main impact estimate is statistically significantly different from zero (e.g., Project NetWork, TETD). These results can be helpful for diagnostic purposes, either for improving certain aspects of the intervention or for considering separate programs for specific subgroups (e.g., for younger beneficiaries).

Despite the inherent variation across studies, there appear to be some broad common patterns across demonstrations beyond practical considerations discussed in the next subsection. In particular, there are recurring differences across age groups (e.g., BOND, MHTS, TETD) and across impairment groups (e.g., BOND, MHTS, Project NetWork, STETS, TETD). SSA has recognized these differences in impacts by age and impairment groups and has conducted demonstrations that focus on specific subgroups of beneficiaries, such as MHTS (and the current Supported Employment Demonstration) for those with impairments related to mental health, or the Youth Transition Demonstration (and the current Promoting Readiness of Minors in SSI demonstration and the Ohio Direct Referral Demonstration) for younger SSI beneficiaries.[20]

There are also some insights arising from findings for specific subgroups in particular demonstrations. For example, the only demonstrations to analyze separate impacts by race/ethnicity were STETS and TETD. In the case of STETS, the results point to a lack of statistical power to isolate subgroup effects for Black non-Hispanic beneficiaries, even though the impact estimate is of similar order of magnitude as other race/ethnic groups analyzed. Other demonstrations have pointed to the potential role of receipt of other (non-SSA) programs, such as "other transfers" (STETS) or Medicaid buy-in (BOPD).

Finally, in several of the demonstrations, there appears to be variation in the treatment effects across program sites participating in the evaluations. This was true for STETS and TETD, but also for BOND and the other demonstrations focused on SSDI or concurrent beneficiaries. Though this variation typically poses some challenges in interpretation, it is not uncommon in large social demonstrations, and some guidance for data collection and analysis has emerged from the past literature on social experiments (e.g., Rothstein and von Wachter 2017).

---

[20] See Chapters 5 and 6 for more information on those.

Practical Considerations from Subgroup Impact Estimates

In terms of practical considerations arising from the discussion of subgroup impact estimates in the earlier subsection "Demonstrations Focused on SSDI and Concurrent Beneficiaries," it might be worthwhile to consider a broad set of common standards for the definition of subgroups, the choice of outcomes, and statistical specifications.

Clearly, the observed differences in outcomes and subgroups between demonstrations shown in Exhibits 7.2 and 7.3 arise partly from differences in the interventions and populations studied. For example, a demonstration focused on, say, beneficiaries who are younger (such as STETS) or beneficiaries with mental health impairment (such as MHTS) might benefit from subgroup definitions different from those for a demonstration focused on a broader population (such as BOND).

Yet, some of the variation across studies appears relatively minor. For that, it might be worth settling on a template of default choices for certain common subgroups that studies can use as a benchmark and modify as needed. To obtain a broader and more consistent coverage of subgroups, such a template could also be used to signal which subgroup categories should be considered for inclusion in future demonstrations (e.g., such as race/ethnicity and gender).

Similarly, the variation in outcomes used appears to be partly due to choices made in the study (e.g., focus on short- versus longer-term earnings outcomes), and partly due to data limitations. For example, BOND and Project NetWork, among others, both use SSA's Master Earnings File to estimate total earnings 2011–2015 and average annual earnings, respectively. The BOND report notes that these data are available only by calendar year and are not precisely aligned with randomization (conducted in May 2011), potentially inflating estimated earnings. In contrast, MHTS estimates earnings using survey data. In future demonstrations, it might be valuable to institute a common set of earnings and benefit metrics based on SSA administrative records available in the same fashion to all demonstrations. To complement information from administrative records while maintaining comparability between studies and between different earnings measures, it is also worth considering a standardized template for survey-based earnings and income measures.

The majority of studies reviewed in this chapter did not systematically report information that would allow us to assess whether subgroup-specific impacts are statistically different from one another within a category (e.g., men's versus women's impacts, within gender). It is worth considering requiring reporting of such tests for all subgroup impacts for future demonstrations.[21] Another aspect is that subgroup estimates come from separate, subgroup-specific treatment-control comparisons. If subgroup characteristics are correlated in the population, we risk attributing subgroup

---

[21] More-recent demonstrations seem more likely to include relevant information, so there may be some acknowledgment of the need for this already. However, it does not appear to be consistent and it could be useful to standardize the reporting of these estimates.

effects to one subgroup because of correlation with another subgroup. For example, suppose that both older beneficiaries and less-educated beneficiaries are found to be less responsive to a benefit offset. Though it might be that both characteristics matter independently in determining the effect of the intervention, it could also be that the effect for older beneficiaries arises if most older beneficiaries are less educated. Similarly, it could be that a subgroup appears to matter only because it is correlated with another subgroup. For example, older individuals tend to have more labor force experience than do younger individuals. Hence, it might be that only experience matters for the outcome of an intervention, but that the age effect when examined alone is found to be statistically significant because of the correlation of age with experience.

A single regression model in which the treatment indicator is interacted with each relevant subgroup category would be able to show the impact of potential correlations among subgroups. The interpretation of the coefficients and their standard errors depends on the specification of the model. For example, *if a main effect for the treatment is included*, one has to exclude one subgroup-treatment interaction from each category.[22] In that case, the main effect for the treatment measures the impact for the excluded category. The coefficients on the subgroup indicators measure the *difference* in the effect of the treatment relative to the excluded group (whose effect is measured by the main treatment indicator), netting out any effects arising from correlation with other subgroups.

For example, if we have three categories—say, gender, a binary age-group indicator (e.g., older versus younger), and a binary indicator for recent work experience (e.g., worked in past five years versus did not work in past five years)—and we excluded the subgroup-treatment indicators for men, younger workers, and no recent work experience, then the main treatment indicator would capture the effect for younger male beneficiaries with no recent employment (the excluded group). The coefficient on the interaction between the treatment effect and an indicator for older beneficiaries would show how the effect for older beneficiaries differs on average from the excluded group, *holding constant* differences in the effect of treatment that could arise because gender and work experience are correlated with age (e.g., if older beneficiaries were more likely to be men and have more work experience). If the subgroup effect for age, considered alone, mattered only because of the correlation of age with recent employment experience, then the coefficient in the interacted model should not be statistically significantly different from zero. In other words, conveniently the standard errors on the interaction effects in the model with multiple interactions can be used to construct test statistics for assessing whether a particular subgroup-specific impact estimate is different from the main effect, *conditional* on inclusion of the remaining subgroup interactions.

---

[22] This assumes, as is commonly the case, that subgroups within a category completely describe the population (e.g., education less than high school, equal to a high school diploma, or more than a high school diploma), such that the indicators for the subgroups add up to the constant term.

If *we do not include the main effect*, then we can include subgroup-treatment interactions for all subgroups to be analyzed. The coefficient on each subgroup-treatment indicator then measures the effect of the treatment on the particular subgroup, holding constant differences in the effect of treatment arising due to correlation with other subgroups. Continuing our stylized example above, if the treatment effect on older individuals is found to be statistically significantly different from zero in this model, then older workers experience a different treatment impact than younger workers, even holding constant the level of work experience.

RECENT ADVANCES IN ESTIMATING HETEROGENEOUS TREATMENT EFFECTS

Clearly, for large populations such as SSDI or SSI participants, the likelihood is high that there is heterogeneity in the response to the particular intervention. Though under random sampling the main treatment effects yield the average treatment effect (ATE) in the relevant population, the intervention might be working better for some groups within that population than for others. For example, this was the case for younger individuals in MHTS or for less-educated workers in Stage 2 of BOND. This section first summarizes existing and new statistical approaches to uncover treatment effect heterogeneity. Then it discusses under what circumstances estimates of treatment effect heterogeneity could be used in future evaluations.

Statistical Approaches to Estimate Extent of Treatment Effect Heterogeneity

The traditional approach of assessing treatment effect heterogeneity in evaluation research is to pursue a limited number of group-level contrasts that are pre-specified in the evaluation's analysis plan. Pre-specification solves the potential bias that could arise if researchers wait to choose contrasts based on the observed outcomes of the evaluation. Pursuing a limited number of contrasts also avoids the risk of finding statistically significant contrasts purely by chance. The traditional approach can yield important insights into treatment effect heterogeneity among key groups relevant to the particular program (see also the discussion in the earlier subsection "Other SSDI-Related Demonstrations"). Yet, by limiting the analysis of heterogeneity to a handful of covariates, the traditional approach might not be able to effectively isolate the relevant margins among which heterogeneity could occur. Indeed, the preceding section, "Discussion of Heterogeneity in Estimates for Demonstration Outcomes," has shown that analyses of even a limited number of subgroups can result in a large number

of potentially imprecisely estimated contrasts even in demonstrations with comparatively very large sample sizes such as BOND.[23]

A growing literature in statistics and economics has devised a range of approaches tailored to the scenario in which the effect of treatments is heterogeneous across individuals. Here I will briefly discuss two broad categories of approaches, one that involves modifying the experiment ex-ante, and another one that affects the way the experimental data are used ex-post.

These two approaches build on important developments in statistics that clarified the assumptions needed and the interpretation of impact estimates in an environment with heterogeneous treatment effects and non-compliance (i.e., when not all individuals take up an offered treatment). Similar considerations apply in an environment where individuals are asked to volunteer for a program, as is the case for the SSA demonstrations. Though a detailed summary of these developments goes beyond the scope of this chapter, the key insight was that individuals could self-select into the program based on their perceived valuation of the treatment, with this self-selection often taken to be the relevant treatment effect (e.g., Angrist, Imbens, and Rubin 1996; Frangakis and Rubin 2002). This insight implies that researchers must make assumptions about individuals' choices after they are exposed to the intervention studied, and it motivates the role of the probability of take-up (the propensity score) discussed below. The literature has shown that a judicious use of such assumptions can yield insights into the nature of heterogeneous treatment effects, and in some cases into the nature of the treatment itself.[24]

These approaches are an important part of the experimental researcher's tool kit and can be helpful in understanding core dimensions of heterogeneity in treatment effects. Their application does vary on a case-by-case basis, they are not meant to deliver fine-grained estimates of individual or group-specific treatment effects, and these approaches are now well-covered elsewhere (e.g., Imbens and Rubin 2015). Ultimately, it is an important empirical question whether the recent methods based on innovations in data science and machine learnings discussed here yield substantive empirical improvements to the standard approach to subgroup analysis or those arising from a deeper understanding program choice.

---

[23] Another potential drawback of standard subgroup analysis is that due to the smaller sample sizes in the subgroups, such analyses are more likely to be capable only of finding evidence of statistically significant effects that are larger than the main effect, rather than vice versa. (Given that significance tests rely on a comparison of an estimate relative to its standard error, and smaller sample sizes imply on average larger standard errors, only larger effect sizes will be found to be statistically significantly different from zero in subgroup analysis.) Yet, for diagnostic purposes, cases in which treatment effects are smaller in absolute value than the main impact estimates are important, as well.

[24] Page et al. (2015) summarize the "principal stratification" approach, as it is called in the statistics literature; and Peck (2013) summarizes the "endogenous subgroup analysis" approach, as it is called in the program evaluation literature.

Ex-Ante Approaches to Improve Estimates of Treatment Effect Heterogeneity

Rothstein and von Wachter (2017) discuss a variety of ways in which experiments could be structured to uncover differences in treatment effects along particular dimensions. Suppose, for example, a continuous measure was available of beneficiaries' (potentially unobservable) underlying earnings potential. For example, the average of an individual's prior earnings is often a strong predictor of future earnings.[25] Then randomization could occur within stratified groups based on predicted earnings potential. Such an approach would, as before, allow estimation of the ATE without a loss in power. In addition, if the sample sizes in each stratum on that measure are appropriately chosen, it would allow a more targeted and interpretable analysis of heterogeneity in treatment effects by differences in labor supply potential.[26] For example, the National Income Tax experiments were stratified based on prior earnings, which can be seen as a predictor for the responsiveness to different income tax rates (Ashenfelter and Plant 1990). An additional advantage of this approach is that it is likely to raise the probability of detecting treatment effects of inducements to return to employment if individuals with the highest estimated earnings potential are also most responsive to such inducements.[27]

Another promising candidate for such cross-classified experiments is a measure of the probability of taking up the offered treatment. In experiments when not all members of the treatment group take up the treatment, research has shown that under some conditions the probability to take up treatment (sometimes called the propensity score) can be used as an index of an individual's benefit from the treatment. Intuitively, if those individuals who benefit more from the treatment are more likely to take up the

---

[25] Alternatively, earnings could be predicted based on information contained in SSA's records on the beneficiary, such as the Residual Functional Capacity questionnaire, or on information from the continuing disability reviews.

[26] Without changing the overall sample size, stratifying the random assignment ensures that sample sizes of strata do not differ due to random variation.

[27] If the effect of treatment does not vary with estimated earnings potential, there is no gain from the stratification in learning about the distribution of underlying treatment effects. However, the resulting subgroups could be of interest for other reasons. For example, the effect of the treatment for low-earning individuals might be of interest in its own right. Moreover, there is also no loss, in the sense that we can still obtain the estimated ATE for all treated individuals.

treatment, then an estimate of the propensity score can be used to isolate those individuals more likely to benefit from the treatment.[28]

Hence, we could design an experiment that cross-classified the treatment using strata defined based on an estimate of the propensity score based on data for the experiment. Again, if sample sizes are kept unchanged, cross-classified randomization ensures balance in sample sizes. An increase in sample sizes for each or at least some strata might be warranted based on standard power calculations. The resulting treatment effects for each stratum would characterize differences in the underlying treatment effect in the population. If stratification by the propensity of take-up is further pursued *within* groups defined by observable characteristics, then in principle the entire distribution of treatment effects for the population can be estimated (e.g., Heckman and Vytlacil 2005).

A closely related approach would be to directly manipulate the probability of take-up of the program with an additional, cross-classified *treatment*. For example, suppose preliminary research showed that distance to the job training or transitional job services site was an impediment to treatment, and that transportation subsidies can increase participation. Then we could cross-classify the randomization of the original treatment (e.g., job training) with randomly assigned transportation subsidies of different amounts. Because the staggered transportation subsidies directly manipulate the probability of take-up, the resulting data can be used not only to better estimate heterogeneity in treatment effect, but also to better understand the effectiveness of approaches for reducing barriers to program take-up.[29]

---

[28] A key assumption is that individuals' choices to take up treatment can be represented by a single summary measure (such as their net monetary gain based on expected increase in earnings minus travel and child care costs). If this is not the case—for example, if individuals with high potential treatment effects are also those who do not understand the benefits from the program—or if the gross gain is of interest to the policymaker (e.g., the effect of treatment on earnings independent of child care or travel costs), then the use of the propensity score as an index of underlying treatment effects has to be reconsidered and potentially modified.

[29] Rothstein and von Wachter (2017) discuss another version of a cross-classified experiment that solves another common problem in the analysis of workforce training. Though often we would like to understand the program impact on an outcome that is observed *conditional* on working (e.g., hourly wages, or annual earnings conditional on working, as in MHTS), the impact estimates on such conditional outcomes are not identified in the basic experimental design because they rely on an endogenous choice—employment—that introduces a potential bias in conditional estimates. Consider then a cross-classified experiment that in addition to the treatment (e.g., a training program) manipulates the decision to work after completion of treatment (e.g., through randomized provision of commuting subsidies or some other relevant means to improve access to job sites). By manipulating the employment decision, this approach allows estimating the effect of training on hourly wages, which is often used as a measure of productivity (or on earnings conditional on working). As a result, the combined experiment allows estimating the effect of the original treatment both on employment and on wages. This is more informative than analyzing annual or even weekly earnings, because earnings reflect both labor supply and wages.

In practice, researchers have to choose how to measure the variable used for cross-classified randomization, such as earnings potential or probability of program take-up. Obtaining such a measure is a key step of the research design. Some of the approaches discussed in the next subsection, "Ex-Post Approaches to Improve Estimates of Treatment Effect Heterogeneity," can be applied to correlational data or to the quasi-experimental studies described earlier in the section "Background on Variation in Employment Potential" to obtain improved measures of earnings potential. Thereby, results from existing demonstrations could be used to guide high-level modeling choices. For example, age, prior employment, or impairment types have shown to be important predictors of employment potential in both quasi-experimental studies and demonstrations; in principle, they could be used to form a coarse ex-ante classification. Alternatively, they could be used to define higher-level groups within which a statistical algorithm provides refinements. Thereby, detailed information available at SSA, such as long earnings histories, medical determination, residual functional capacity, and place of residence, could be combined with information on indicators of local labor markets, such as the incidence of vacancies in jobs similar to the occupation previously held by a beneficiary.

From a theoretical point of view, the most accurate possible measure would be preferable, but as further discussed later (in "Practical Considerations for Use of Heterogeneous Treatment Effect Estimates"), practical considerations could lead researchers to choose variables that can be potentially observed as the program is administered by case managers. For example, the earnings potential of SSDI beneficiaries could be based on the amount of prior earnings, on their predicted earnings (e.g., based on their demographics, education, occupation, and employment experience), or on a more sophisticated measure based on an assessment of the market value of their residual functional capacity or their occupation-specific skills.[30] Similarly, the estimated probability of treatment take-up could be based on the full set of information available to SSA, on a subset that is deemed sufficiently predictive of take-up, or on additional variables that currently might not be routinely collected but were found to be predictive of program take-up in preliminary studies for the particular experiment or in analysis of data from existing demonstrations.

---

[30] As an alternative, we could use estimates of heterogeneity in labor supply effects to SSDI benefits from studies that seek to use random variation occurring naturally in the data as discussed earlier in "Discussion of Heterogeneity in Estimates for Demonstration Outcomes" (e.g., French and Song 2014; Hemmeter and Bailey 2016; Maestas, Mullen, and Strand 2013). Though preferable insofar as these are already estimated treatment effects from being exposed to the SSDI program in various ways (depending on the study), they might not be as suitable for diagnostic or targeting purposes because these treatment effects cannot be easily measured in the population (see the subsection "Practical Considerations for Use of Heterogeneous Treatment Effect Estimates").

Ex-Post Approaches to Improve Estimates of Treatment Effect Heterogeneity

The potential statistical issues of traditional approaches to subgroup analysis have been a motivation for algorithmic (or data-driven) approaches to exploring treatment effect heterogeneity that can be applied to existing experimental data. Several approaches have been developed that implement statistical algorithms by which the computer, not the researcher, runs through a large number of contrasts based on flexibly defined categorical groups. From that, it obtains estimates of treatment effects for these contrasts and their standard errors.

The following gives a broad overview of two groups of recent approaches.[31] Based solely on the way the treatment effect is calculated, I will refer to these as "semi-parametric" and "non-parametric" approaches. All of these approaches are designed to explore treatment effect heterogeneity among a large number of categories (e.g., age by gender by education by income class). They differ in their data requirements and in their statistical properties. Depending on the particular application, they can differ in the interpretability of the resulting subgroups, as well.

Semi-parametric. What I call "semi-parametric" approaches can be viewed as an extension of the traditional approach. As discussed regarding the traditional approach in the opening of this "Statistical Approaches" section, a semi-parametric approach can be formulated in a standard linear regression framework in which the desired outcome is the outcome variable, and the treatment dummy interacted with the pre-specified subgroup indicators are the main control variables.[32] The coefficients on the interactions represent the group-specific treatment effects (or differences in the effects with respect to an omitted category, if a constant term is included).[33] Recent publications have extended this approach to allow estimation of treatment effects that are a fully flexible function of all the covariates, rather than just a function of group indicators. Various statistical approaches are used to automatically choose the most relevant treatment effect differences among a potentially large number of contrasts. Because no restriction is placed on how the treatment effects depend on the observed covariates, this approach can be viewed as non-parametric estimation of how treatment effects vary with covariates.

A version of this semi-parametric approach shows intuitively how estimates of individual-level treatment effects can be modeled flexibly as a function of covariates.

---

[31] See Peck (2005) for an example of an early approach to data-driven choices of subgroup contrasts that uses statistical methods to obtain a limited number of clusters of similar observations in the population that can be used instead of or in addition to traditional groups for obtaining subgroup contrasts.

[32] In the case of an experimental evaluation, additional control variables are sometimes added for precision, but I ignore these here for simplicity.

[33] Without a constant term, the coefficients simply measure the treatment effects by group. Alternatively, if the coefficients on interactions are constrained to sum to zero, they represent the difference with respect to the mean treatment effect.

Consider drawing for each treated subject a statistical twin from the pool of control participants. This can be done, for example, by choosing two individuals who have the same estimated probability of treatment (the propensity score discussed earlier in the "Ex-ante Approaches" subsection). The difference in outcomes between these two individuals can be viewed as a coarse estimate of the individual treatment effect. This individual treatment effect can then be regressed on a flexible functional form of the covariates, as in any standard semi-parametric or non-parametric regression. Standard machine learning and related approaches (e.g., least absolute shrinkage and selection operator, or LASSO) can be used to choose among a potentially large number of covariates.[34]

The resulting estimates can be used to predict a treatment effect for each individual based on their covariates' values (something also referred to as conditional ATE, or CATE, because it is conditional on the given individual's observed covariates). The distribution of estimated CATEs can be analyzed as a whole or for different subgroups to assess the estimated degree of heterogeneity among treatment effects. The CATEs can also be used, in principle, to re-calculate the overall treatment effect, as if different populations of individuals had received treatment (e.g., under different program rules or different outreach strategies).[35] Such an exercise can be a useful diagnostic device to assess how much potential room for improvement there might be for better targeting the intervention.

Non-parametric. The second, non-parametric approach dispenses with the linear regression framework and uses the patterns in the data directly to isolate those groups that have the largest differences in treatment effects. This is done by recursively partitioning the data into those groups that exhibit the largest difference in treatment effects. The algorithms used for the stepwise partitioning (called "causal trees") can more flexibly search over different combinations in the data than the semi-parametric approach can, and hence might be more likely to isolate salient differences in treatment effects. The final product is again an estimate of a CATE for each individual based on their covariates. These estimates can be used for analysis of treatment effect heterogeneity in the same way as the results from semi-parametric estimation can.[36]

---

[34] The same result can be achieved by separately modeling the counterfactual outcomes under treatment and non-treatment as a function of covariates, and then constructing individual treatment effects as a difference of each individual's estimated counterfactual outcomes (e.g., Foster, Taylor, and Ruberg 2011).

[35] This is pursued, for example, by Knaus, Lechner, and Strittmatter (2020), who use the semi-parametric approach to estimate the effect of job search programs in Switzerland.

[36] The CATEs obtained from this method have been shown to have desirable statistical properties (Athey and Imbens 2016).

Choosing a Statistical Approach

Different approaches differ in their data requirements, ease of implementation, and statistical properties. For example, the semi-parametric approaches can be used to choose among a large number of covariates with respect to the total sample size. The non-parametric approaches discussed can also deal with a large number of covariates but tend to require a larger number of observations. In both cases, a higher number of observations and a higher coverage of observations along possible dimensions of heterogeneity will lead to more accurate estimates of the underlying heterogeneity in treatment effects.

Though the algorithms replace the researchers' potentially confounding choices of dimensions of heterogeneity, it is important to bear in mind that researchers will still have to specify aspects of the analysis that can influence the final outcome (e.g., the smoothing parameters of the machine learning algorithms, or the partitioning of the data into training and estimation samples). As this again risks introducing researcher-induced variation in outcomes, sensitivity analyses along the relevant margins are an important step in the implementation of these measures.

Another feature of all of the statistical approaches mentioned is that the result might or might not be readily interpretable. Interpretability might not be required if, for example, the main goal is to predict which individuals will most benefit from treatment to better assign treatment directly based on the estimated CATEs. However, as further discussed next, in some cases, interpretability is a desired feature of the analysis, for example, if the analyses are meant to inform the understanding of the treatment more broadly or if the results are meant to be used to generate new assignment mechanisms. However, the researcher could modify the type of covariates used for the heterogeneity analysis and see whether a reduced set of covariates is still able to provide a good fit to the observed heterogeneity in treatment outcomes.

Practical Considerations for Use of Heterogeneous Treatment Effect Estimates

Several potential practical and ethical aspects arise when considering the potential use of estimated heterogeneous treatment effects in practice. In the following, I will briefly discuss such aspects for three potential use cases for estimated heterogeneous treatment effects: diagnostic purposes, intervention targeting, and intervention evaluation.

For Diagnostic Purposes

An inspection of the estimated CATE discussed above in the "Statistical Approaches" section can indicate for which individuals the intervention might not be working as well as for others. Systematic pattern in the CATEs might further give clues as to the underlying sources of the differences. One important question that arises is what constitutes the underlying sources for heterogeneous outcomes. In the context

of the demonstrations discussed in this chapter, treatment effect heterogeneity is likely to arise from individual characteristics that affect the ability or desire to work. These characteristics could lead to differences in the effect of the treatment, in the sense that as designed, the treatment does not "work," or works less for some individuals than for others.

Yet, treatment effect heterogeneity could also arise from other sources than the nature of individuals and of the treatment. It has often been observed that different program sites exhibit different average treatment outcomes. Such "site effects" can arise due to differences in how the program is implemented or administered, which in turn can influence the composition of individuals being served.[37] Heterogeneity in outcome could also arise from differences in participants' understanding of the intervention, interactions with caseworkers or teachers, or access to other employment supports, just to name a few.

The analysis of the estimated CATEs can be useful in this process, as different variables used for estimation can reflect some of the potential underlying sources of heterogeneity. Follow-up research (e.g., quantitative or qualitative surveys, focus groups, or reviews of program fidelity and program process, among others) is likely needed to complement the quantitative analysis. Besides being directly useful to the problem at hand, this information can also help to provide information on what data can be collected in future demonstrations to obtain more informative estimates of CATE.

For Intervention Targeting

Heterogeneous treatment effects can also in principle be used for better targeting the intervention to those individuals who will benefit the most. Whether this is desirable ultimately could be a decision based on a range of factors outside of the realm of quantitative analysis. However, in principle, the estimated CATEs, possibly together with estimates of the cost of treating individual participants, could be used to generate lists of individuals for whom the estimated cost-benefit of the intervention is particularly high. Such a list could be prioritized for treatment or for proactive outreach to take up treatment. As further discussed below, when pursuing such a strategy, it is important to consider potential risk of propagating pre-existing biases in program access or success.

From a practical point of view, this targeting requires that the information used to calculate the CATE be available to administer the program on daily basis. For example, if mostly based on administrative data that are accessible in real time, the CATE could be calculated based on up-to-date information for use by a caseworker or an outreach team. In other instances, the CATE might be too complex; that is, based

---

[37] The composition of individuals can differ across sites for other reasons (e.g., differences in local populations), and controlling for such differences when evaluating site effects can be a useful step in diagnosing potential sources of site differences.

on information not easily accessible to SSA (e.g., if it was based on survey data collected for a demonstration) or based on administrative data not readily available to caseworkers on the ground. In this not uncommon scenario, the research team can assess whether it is feasible to generate informative estimates of the CATE based on a subset of variables that are more readily available to relevant caseworkers or that could be collected at low cost as part of a modified intake process (e.g., in the form of a short intake or targeting questionnaire).

A potential concern for targeting is that the estimated CATEs are subject to sampling error. If a fixed cutoff for a CATE were to be used for prioritizing individuals for treatment, this would lead both to false negatives (lower prioritization of individuals whose true CATE is above the cutoff) and to false positives (higher prioritization of individuals whose true CATE is below the cutoff). Depending on estimates of the size of the sampling error and a given cost of making an error, the cutoff could be sufficiently relaxed to avoid making mistakes that are deemed too costly. More generally, given variability from sampling, together with normal uncertainty regarding the correct statistical model or the CATEs, appropriate caution is advised when using estimated CATEs to exclude individuals from treatment outright.[38]

Another concern is that estimated CATEs inadvertently propagate pre-existing biases or discrimination. For example, if in the past no effect was found on a subgroup because of the influence of racial or ethnic discrimination, then using the resulting CATEs into future targeting would risk propagating that same discrimination. This is a well-known problem in the literature using machine learning algorithms to predict future outcomes. Such "predictive analytics" can be useful in its own right for targeting by helping to identify which individuals are more likely to take up an intervention. Predictive analytics also can be used to generate probabilities that are useful for stratified analysis (e.g., predicting the probability of a particular beneficiary working above the SGA level). Approaches used in the literature on predictive analytics for detecting bias, together with institutional and qualitative information on the intervention studied, can be used to prevent propagating any biases in treatment assignment that might be introduced by the use of estimated CATEs.

For Intervention Evaluation

Finally, the estimated CATEs could be used to inform the overall evaluation of the viability of a tested intervention. Though not the sole decisive factor, commonly estimated ATEs of an intervention play an important role in its evaluation. The

---

[38] There could be cases in which such an approach is reasonable. Consider the case when the expected effect of an intervention is positive. The treatment might not be considered viable for individuals with precisely estimated but large and negative CATEs, or precisely estimated but low and positive CATEs *and* high program costs.

presence of treatment effect heterogeneity makes matters more complicated, but the CATEs can provide a valuable source of information.

Consider the case of a single outcome of interest—say, the net impact of the program on individuals' disposable income (i.e., earnings plus SSDI/SSI benefits). We can represent society's valuation of the intervention as a weighted sum of treatment effects among all participants (this is sometimes called a "welfare function"), where the (welfare) weights represent the value to society of providing additional income to each individual. In this basic, yet realistic example, if society's values are equal for all individuals, heterogeneous treatment effects do not provide additional information about the program beyond the ATE. Yet, this would be a very unusual welfare function. In the United States, and in many other countries, programs are structured such that funds are transferred to lower-income individuals, reflecting that welfare weights for many social insurance programs usually decrease with income.

Suppose then that the distribution of estimated treatment effects is centered around zero impact, such that the ATE is zero as well, but that treatment effects are positive for poorer or otherwise needier individuals. In the realistic scenario of welfare weights that decrease with income, all else equal, the evaluation of such a program would be very different under ATE and CATE. The question can become more complex with multiple primary outcomes of interest. For example, in all demonstrations discussed here, measures both of earnings and of benefit receipt were primary outcomes. In this case, the estimated CATE on net income would likely receive a different weight (the welfare weight of program participants) than would the estimate of total program costs saved, which affects workers paying Social Security taxes (and hence would be evaluated at the welfare weights of, say, an average worker). Though more complex, this is not an unusual set of welfare tradeoffs, and hence estimated CATEs could be a valuable input into a broader process of evaluating the viability of proposed programs.

CONCLUSIONS

This chapter has discussed the heterogeneity of the impacts evaluated in eight demonstrations that tested the impact of work incentives and work supports for SSDI beneficiaries and SSI recipients. An analysis of heterogeneous impacts allows a better understanding of whether some participants benefit more from a particular intervention than other participants do. This, in turn, allows SSA to either target potentially expensive interventions to individuals who will benefit most or improve interventions for those participant subgroups that appear to benefit less from it.

The chapter started out by motivating the potential role of heterogeneity in beneficiaries' and recipients' employment potential, based on descriptive and non-experimental evaluations. The chapter then provided an overview of findings on heterogeneity, including a summary of the comparability of outcomes and subgroup definitions among demonstrations. It then briefly discussed potential lessons and practical implications from the discussion of heterogeneity. It then concluded with a

summary of alternative statistical approaches to address heterogeneity in the response to treatments brought forward in the recent literature, distinguishing between approaches modifying the experimental design versus those based on existing experimental data.

The chapter comes to four broad conclusions.

1. **The available results of subgroups paint a helpful but complex picture of heterogeneity of impact estimates.**

For MHTS, one of the two demonstrations with a detailed analysis of impact estimates by subgroup, the intervention (improved behavioral health services and case management) led to widespread increases in earnings for all subgroups studied, with the exception of younger beneficiaries (younger than age 35). These findings (and findings from the other demonstrations discussed in this chapter) are based on study participants drawn from a pool of volunteers and hence might not be representative for all beneficiaries with these health conditions. Yet, as long as those volunteering to participate in the demonstration are similar to those who would take up the treatment, were it offered on a larger scale, the estimated impacts are informative if the program were adopted more widely.

For BOND, the other demonstration with a systematic analysis of subgroup impacts, the subgroup analysis identifies several groups of Stage 2 volunteers for whom the treatment (benefit offset) increased earnings. This is true for younger and less-educated beneficiaries, those with some prior unemployment (at a marginal level of significance), and those with "other impairments" (other than major affective disorder or a musculoskeletal disorder, the two impairment types studied explicitly in BOND). This is notable because the overall effect on earnings was found not to be statistically different from zero. In contrast to MHTS, the BOND report shows tests for the difference within subgroups. Only the contrast between individuals with prior employment and prior unemployment almost satisfies the margin of being statistically significantly different from zero at a 10 percent level.

2. **It would be useful to improve comparability of estimated program impacts between demonstrations by adopting a core set of common definitions of subgroups and outcomes.**

Currently, SSA's demonstrations used a range of definitions of earnings and employment, making comparison between studies difficult. Similarly, demonstrations used a broad range of different subgroups, with only age, enrollment type (e.g., SSDI versus concurrent SSDI/SSI), and to some extent impairment type being comparable across studies. Settling on comparable earnings and employment measures, harmonizing subgroup definitions, and including gender, race/ethnicity, and prior education in standard subgroup analysis would be worth considering. In terms of implementation, a common set of reporting practices would be helpful (e.g., reporting of test statistics for differences of subgroup estimates).

3.   **It would be useful to harmonize statistical analysis of demonstrations by reporting results from statistical tests of the difference between subgroup impacts and estimate statistical models that account for cross-group correlations.**

The review of empirical findings in this chapter suggests the lessons learned from SSA's demonstrations could be further improved by adopting a common set of standards for the statistical analysis of impact estimates by subgroups and the reporting of its results. All analyses should report statistical tests for the differences in subgroup impacts within a category (e.g., men vs. women). In addition to comparisons of differences in treatment and control group means for each subgroup, subgroup impacts should also be estimated from statistical models that include interactions of the treatment indicator with all relevant subgroups to account for potential correlations among groups in the population.

4.   **It is worth exploring new ex-ante and ex-post statistical approaches to analyzing impact heterogeneity.**

Such approaches could allow analysts to assess impact heterogeneity for existing demonstrations using newly developed data science techniques. For example, in future demonstrations, stratifying randomization by comprehensive measures of work potential and oversampling some strata could improve the ability to detect impacts. Similarly, data science methods have been used to estimate treatment effects after randomization that flexibly use patterns in the data to determine relevant subgroup differences. Such approaches could lead to a richer understanding of how proposed programs affect the large and heterogeneous population of SSDI beneficiaries and SSI recipients and could allow improved service outcomes through improved targeting or intervention differentiation.

Consider the example of a demonstration of a potential new work support intervention, or an existing program that has not yet been experimentally evaluated, such as Ticket to Work. It is possible to develop approaches to target Ticket to Work, based on econometric estimates of employment potential or of the probability of taking up the program, using machine learning approaches. As discussed in this chapter, it could be possible to further improve such targeting and resulting intervention outcomes by experimentally estimating heterogeneous treatment effects. To maximize the ability to identify relevant heterogeneity in treatment effects, such a demonstration would settle, in advance, on a stratification within which randomization takes place (such as predicted employment potential based on rich administrative data available at SSA) and adjust sample sizes accordingly. An important preliminary step is research in using SSA's substantial data (containing among them information on earnings histories, medical determination, the occupation, labor market experience, and residual functional capacity) together with growing information on employment and vacancies in local labor markets to improve predictions of beneficiaries' work potential.

In addition, once the experiment has taken place, increasingly standard data science tools can be used to assess further dimensions of treatment heterogeneity and further refine these strata. The resulting distribution of estimated treatment effects (the CATEs) can then be analyzed for diagnostic purposes. Researchers can then use these treatment effects to devise approaches that can be implemented in the field to better target the intervention to those beneficiaries most likely to take up and to benefit from it. If desired, the design of the experiment and data collection can be adjusted in advance to meet the envisioned use case of the information on heterogeneity.

Finally, the review of the mixed success of the eight demonstrations to achieve broad and sustained earnings and reductions of SSDI receipt among SSDI beneficiaries (with evidence of increases in SSDI receipt in BOND) suggests that SSA should be intervening when disabled workers are younger or as soon as a new disability occurs. SSA has already begun assessing the effectiveness of early interventions (see Chapter 5 in this volume). It may be worth further exploring potential synergies between workforce interventions for hard-to-reemploy workers and those that serve partially disabled workers. Aiming to understand how partially disabled workers fare in these workforce programs, and assessing the potential costs and benefits of reintegrating the workers prior to their receipt of or application to SSDI seem worth considering in the future.

NOTE

A version of this chapter with additional detail about the demonstrations and their findings in appendix tables is available online (http://www.econ.ucla.edu/tvwachter/).

Contributor

Till von Wachter, Professor of Economics; Faculty Director, California Policy Lab; Director, Federal Statistical Research Data Center; and Associate Dean for Research, Social Science Division, University of California Los Angeles—Dr. von Wachter's research examines how labor market conditions and institutions affect the well-being of workers and their families.

Chapter 7
Comment

Howard H. Goldman
University of Maryland

My comments (regarding "An Overview of Current Results and New Methods for Estimating Heterogeneous Program Impacts," by Till von Wachter) derive from my perspective, gained over the past 40 years, as a mental health policy researcher and occasional advisor to the Social Security Administration's programs on disability. I address myself to three areas:

- Salient points from this excellent chapter with which I strongly agree.
- A point of disagreement with respect to characterization of mental impairments as less severe and people experiencing them as having greater employment potential than other groups.
- General policy lessons about disability due to mental impairments based on two demonstrations, the completed Mental Health Treatment Study (MHTS) and the Supported Employment Demonstration (SED), which is still in the field.

POINTS OF AGREEMENT

In my view, the most important general lesson on heterogeneity stated in the chapter is this: "Insights on variation in the effects of treatment for certain groups can help in better implementing interventions by informing which SSDI beneficiaries and SSI recipients might be particularly responsive to new features of a program." Recognizing heterogeneity in outcomes is important to focusing an intervention on a particular target group likely to benefit. Such targeting is important both for the effectiveness of the intervention and for recruitment of participants in a demonstration. These observations will be critical, as well, if we ever implement a program based on a demonstration. It is true that no one-size-fits-all intervention is likely to emerge, but it is still possible to expand the scope of an intervention to focus on a broader group of potential participants. For example, supported employment that follows the Individual Placement and Support (IPS) model is central to both the MHTS and the SED. Both demonstrations focus on individuals with mental impairments, but the MHTS successfully targeted individuals who receive disability benefits, whereas the SED is focusing on individuals initially denied disability benefits based on their mental impairments and comorbidities. The SED tests the dismantling of the intervention from the MHTS, comparing a Full-Service arm, which includes a nurse care manager, versus a Basic-Service arm without the nurse. It remains to be seen whether the expansion of the interventions to individuals denied benefits on initial application (with either intervention arm) is warranted. Meanwhile, in other studies in the field,

IPS is also being tested on individuals with post-traumatic stress and substance-use disorders. Conceptually, IPS could be tried with individuals with any category of impairment or mix of impairments. These demonstrations illustrate the potential benefits of focusing on the heterogeneity of outcomes.

Another important point made by the author is the conclusion that "groups studied in the literature are quite coarse." The MHTS was targeted on individuals with severe mental impairments, psychotic and affective disorders, whereas the SED is being tried on individuals who allege a much broader array of mental and general medical impairments. The more focused MHTS found differences in impact of IPS and integrated behavioral health treatments on younger beneficiaries with schizophrenia and on older beneficiaries who experienced depression. Older beneficiaries, presumed to have more work experience, fared better in employment outcomes. We will have to wait several years for the results of the impact analysis of the SED on different groups, as that study is not yet completed.

Although coarsely defined in the literature, some groups have more employment potential than others. The chapter points out, however, "even those SSDI beneficiaries with employment potential can face substantial labor market barriers and be at risk of financial hardship absent benefits." In my experience, this is true for individuals with disability due to mental impairments, who face hiring impediments based on prejudicial attitudes toward individuals with mental impairment. The MHTS and SED both illustrate the problems faced by such individuals in obtaining employment.

POINT OF DISAGREEMENT

I want to take exception to the characterization of individuals with mental impairments, particularly younger individuals, as having less severe impairments and thus having higher employment potential. Although this may be true of the findings of several studies reviewed in the chapter, this characterization is not uniformly true. Mental impairments often are invisible, and they tend to wax and wane. An individual with a mental impairment might seem to be able to work one day, and be unable to work on another, establishing a pattern of inconsistent work attendance and lack of productivity that is not conducive to full-time employment. The functional limitations imposed by mental impairments affect the full range of work demands, making any kind of work on a sustained basis a challenge. Younger individuals may have more years of potential for employment, which is a hopeful perspective, but many of the most severe mental disorders have their onset in the late teens and early twenties, and those individuals with earlier onset often are more severely functionally impaired than those with later presentations of their conditions.

The MHTS demonstrated the mixed experience of employment potential for individuals with mental impairments, particularly younger individuals. Although some 60 percent of individuals in the treatment arm of the MHTS had some level of competitive employment, compared with the control group at 40 percent, none stopped

receiving SSA disability benefits. Full-time employment was an elusive goal for this group, which sets up a final comment.

A GENERAL POLICY LESSON FROM THE MHTS AND SED

As noted in the chapter and in my comment above, none of the participants on the MHTS worked above the Substantial Gainful Activity (SGA) threshold of the SSA statutory definition of disability, and none exited SSA's disability programs. Based on my interviews, as a part of the SED evaluation team, I can report that beneficiaries and service providers, alike, are coming to view being on benefits and working some, within the SSA rules, as a good outcome. Participants in both the MHTS and the SED are working, most at less than the SGA level, but they are enjoying some of the benefits of social inclusion and work force participation. They do not achieve the policy goal of savings to SSA, but their mental health has improved, and they are experiencing a higher degree of social integration. SSA is to be applauded for supporting these demonstrations and these broader outcomes. And perhaps we should not be surprised that individuals found disabled by SSA are not able to work above SGA levels, even with substantial supports, because they have been found disabled under a very strict standard of disability.

Howard H. Goldman, Professor of Psychiatry, University of Maryland School of Medicine—For the past decade, Dr. Goldman has chaired the National Academies of Sciences, Engineering, and Medicine, Standing Committee of Medical and Vocational Experts Assisting the Social Security Administration (SSA) on Disability Issues.

Chapter 7
Comment

Nick Hart
The Data Foundation

More than 40 years ago, Lee Cronbach and Associates (1980) wrote that context matters in evaluation. Cronbach's plea for increased consideration of external validity in design and execution of evaluations is a poignant message for evaluators in the 21st Century, including with regards to Social Security Disability Insurance (SSDI) demonstration projects. Shortly after being sworn into office on January 20, President Joe Biden (2021) signed a first Executive Order directing agencies to reimagine how to assess and solve for inequities in government programs. Given this context about the appeals from the evaluation community for 40 years and new policy expectations to understand disparities and inequities, it is striking just how tremendous the gaps are in analyzing key impacts across relevant subgroups in past SSDI demos.

In "An Overview of Current Results and New Methods for Estimating Heterogenous Program Impacts" by Till von Wachter, the author's review of past demonstration projects alongside the strong evaluation infrastructure in place at the Social Security Administration (SSA) leads to a clear conclusion: analysis of beneficiary subgroups can be substantially improved. Some subgroup analysis did occur, as von Wachter notes, so this is not to say subgroup analysis was nonexistent at SSA. Yet only two of seven randomized evaluations analyzed differences by race and ethnicity, for example. The policy implications are vast and the risks for decisionmakers from gaps in knowledge where disparities or inequities may exist are tremendous, particularly when compared to the relatively low cost of increasing sample sizes or adjusting contracts for additional subgroup analytics in a program with $140 billion in benefit payments per year.

THE POLICY CASE FOR ANALYZING HETEROGENOUS IMPACTS

In 2015 when SSDI reforms were considered and negotiated by Congress and the White House, the policymakers turned to existing demonstrations to consider areas for reform, savings, improvement, and further study (McCann and Hart 2019). We know the policymaking community relies on insights from demonstration projects to inform actual policy—this is not theory; this is the practice and tradition for SSDI. In many respects, the historical use of evidence from SSDI demos is also a testament to the quality and effort from SSA and its partners in testing strategies for improving programs and services. Policymakers need relevant insights about subgroups that differentially experience SSDI as well as respond with variation across theorized program improvements. A single average treatment effect for a study population or sample is insufficient.

The major challenge that von Wachter identifies and articulates about the gaps in some demonstration projects of analysis of heterogenous effects is a theme that should have been addressed long ago as part of the design and planning of demonstration projects. In 2021 and beyond, improved analysis that bolsters external validity is an imperative. President Biden's Executive Order is an impetus; so too is the Foundations for Evidence-Based Policymaking Act (Evidence Act) and the ensuing evaluation standards that call for ethical evaluations in government that address "contextual factors that could influence the findings or their use" (Vought 2020, 5). The policy case for analyzing subgroups can be summarized to say there is both an expectation and an ethical obligation for the SSA and its evaluators to conduct analyses that explore heterogenous impacts across relevant subgroups. Doing so supports efforts to understand inequities and enables improved targeting of efficacious interventions to the individuals most likely to realize benefits.

IMPROVING ANALYSIS OF HETEROGENOUS IMPACTS RETROSPECTIVELY AND PROSPECTIVELY

The suggestions offered by von Wachter in his chapter for embedding heterogenous effects are practical, salient, low cost, and necessary. If anything, the major critique of von Wachter's chapter is that it simply does not go far enough in suggesting, given vitality of the topic, that much more should be and must be done by SSA in the future to encourage more subgroup analysis. These actions cannot solely be the responsibility of the research and evaluation community, but must be reflected by SSA leadership and stakeholders.

First, SSA should continue to plan for future evaluations to address heterogenous impacts when possible, with actions that could include these:

- *SSA Evaluation Policy.* In implementing SSA's published evaluation policy required by the Evidence Act and incorporating the required evaluation principles, SSA can explicitly reflect its own policy statement to prioritize particular types of analytics necessary to improve programs for SSDI beneficiaries as a key way to tailor findings to meet the needs of evaluation users (SSA 2020f).
- *SSA Learning Agenda and Equity Assessment.* In complying with the Evidence Act's requirements and the President's Executive Order, SSA can continue to effectively collaborate with its stakeholder communities and policymakers to identify key questions or themes to incorporate in future research and evaluation plans, to specifically study and address inequities or disparities. For example, key questions could include better assessment of perceptions and burdens, challenges in access, or denials of benefits by subgroups, including stratification of results by race and ethnicity when appropriate.

- ***Outcome Standardization.*** In leveraging the Evidence Act, SSA can begin to identify shared definitions and standards for interpreting particular outcome measures to enable improved comparability across studies, evaluations, and demonstration projects.

In addition to planning for the future, SSA can also potentially supplement insights from completed projects by leveraging its own existing data infrastructure to reanalyze past interventions through data linkages or sharing with other agencies and partners. The US Commission on Evidence-Based Policymaking unanimously recommended to the President and Congress in 2017 that federal agencies enhance data sharing and linkage capabilities (CEP 2017). SSA has a compelling case to be on the forefront of these linkage activities moving forward, including to apply current data to past research in order to generate new insights about long-term impacts across subgroups. Related, access to SSA data for retrospective evaluation through de-identified data sets or secure data enclaves could better support SSA's and the disability community's long-term evidence-building needs. Though some such activities may be underway at SSA today, updating procedures, regulations, and notices under the Privacy Act and SSA's authorizing statutes can be time intensive and burdensome, so prioritizing these activities should be a priority for SSA's chief data officer, evaluation officer, and other senior leaders.

CONCLUSION

SSA has a vital role in providing decisionmakers and stakeholders with relevant information about what works best, in what contexts, and for whom. Analysis of group effects in the future is essential for SSDI demonstration projects, and SSA should act upon the insights offered by von Wachter, including exploring new innovative mechanisms and approaches for addressing contemporaneous methodological and resource constraints. SSA must also reimagine its data and evaluation capabilities to ensure appropriate information is available as open data and for other evaluative activities, as well. SSA has historically been a leading federal agency for enabling evidence-informed decisionmaking, but there remains much room for progress and improvement in the years ahead.

Nick Hart, President, The Data Foundation—Dr. Hart leads a Washington, DC-based non-profit think tank that works to improve government and society by encouraging the use of data to improve public policymaking. Dr. Hart has worked on a wide range of issues including Social Security, disability, anti-poverty, environmental, energy, economic development, and criminal justice policies.

Volume References

Abraham, Katharine G., and Melissa S. Kearney. 2020. "Explaining the Decline in the US Employment-to-Population Ratio: A Review of the Evidence." *Journal of Economic Literature* 58 (3): 585–643.

Administration for Community Living. 2020. "Community Integrated Health Networks." https://acl.gov/sites/default/files/common/BA_roundtable_workgroup_paper_20 20-03-01-v3.pdf.

Aizer, Anna, Nora E. Gordon, and Melissa S. Kearney. 2013. *Exploring the Growth of the Child SSI Caseload in the Context of the Broader Policy and Demographic Landscape*. Cambridge, MA: National Bureau of Economic Research.

Almond, Douglas, and Janet Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives* 25 (3): 153–172.

Anderson, Mary A., Gina Livermore, AnnaMaria McCutcheon, Todd Honeycutt, Karen Katz, Joseph Mastrianni, and Jacqueline Kauff. 2018. *Promoting Readiness of Minors in Supplemental Security Income (PROMISE): ASPIRE Process Analysis Report.* Washington, DC: Mathematica Policy Research.

Anderson, Catherine, Ellie Hartman, and D. J. Ralston. 2021. "The Family Empowerment Model: Improving Employment for Youth Receiving Supplemental Security Income." Washington, DC: US Department of Labor, Office of Disability Employment Policy.

Anderson, Catherine A., Amanda Schlegelmilch, and Ellie Hartman. 2019. "Wisconsin PROMISE Cost-Benefit Analysis and Sustainability Framework." *Journal of Vocational Rehabilitation* 51 (2): 253–261.

Anderson, Michael, Yonatan Ben-Shalom, David Stapleton, and David Wittenburg. 2020. *The RETAIN Demonstration: Practical Implications of State Variation in SSDI Entry.* Report for Social Security Administration. Washington, DC: Mathematica Policy Research.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–455.

Arnold Ventures. 2020, December 15. "National RCT of 'Year Up' Program Finds Major, Five-Year Earnings Gains for Low-Income, Minority Young Adults." Straight Talk on Evidence. https://www.straighttalkonevidence.org/ 2020/12/15/national-rct-of-year-up-program-finds-major-five-year-earnings-gains-for-low-income-minority-young-adults/.

Ashenfelter, O., and M. W. Plant. 1990. "Nonparametric Estimates of the Labor-Supply Effects of Negative Income Tax Programs." *Journal of Labor Economics* 8 (1): S396-S415.

Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *Proceedings of the National Academy of Sciences* 113 (27): 7353–7360.

Autor, David H., and Mark G. Duggan. 2000. "The Rise in Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics* 118 (1): 157–205.

Autor, David H., and Mark G. Duggan. 2006. "The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding." *Journal of Economic Perspectives* 20 (3): 71–96.

Autor, David, H., and Mark G. Duggan. 2007. "Distinguishing Income from Substitution Effects in Disability Insurance." *American Economic Review* 97 (2): 119–124.

Autor, David H., and Mark Duggan. 2010. *Supporting Work: A Proposal for Modernizing the US Disability Insurance System*. Washington, DC: Center for American Progress and the Hamilton Project.

Autor, David H., Mark G. Duggan, Kyle Greenberg, and David S Lyle. 2016. "The Impact of Disability Benefits on Labor Supply: Evidence from the VA's Disability Compensation Program." *American Economic Journal: Applied Economics* 8 (3): 31–68.

Autor, David H., Nicole Maestas, Kathleen J. Mullen, and Alexander Strand. 2015. *Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability Applicants*. Cambridge, MA: National Bureau of Economic Research.

Autor, David, Nicole Maestas, and Richard Woodberry. 2020. "Disability Policy, Program Enrollment, Work, and Well-Being among People with Disabilities." *Social Security Bulletin* 80 (1): 57.

Bailey, Michelle Stegman, Debra Goetz Engler, and Jeffrey Hemmeter. 2016. "Homeless with Schizophrenia Presumptive Disability Pilot Evaluation." *Social Security Bulletin* 76 (1): 1–25.

Bailey, Michelle Stegman, and Jeffrey Hemmeter. 2015. "Characteristics of Noninstitutionalized DI and SSI Program Participants, 2013 Update." *Social Security Administration Research and Statistics Notes*. No. 2015-02. Social Security Administration. https://www.ssa.gov/policy/docs/rsnotes/rsn2015-02.html.

Bailey, Michelle Stegman, and Robert R. Weathers II. 2014. "The Accelerated Benefits Demonstration: Impacts on Employment of Disability Insurance Beneficiaries." *American Economic Review: Papers & Proceedings* 104 (5): 336–341.

Baller, Julia B., Crystal R. Blyler, Svetlana Bronnikov, Haiyi Xie, Gary R. Bond, Kai Filion, and Thomas Hale. 2020. "Long-Term Follow-up of a Randomized Trial of Supported Employment for SSDI Beneficiaries with Mental Illness." *Psychiatric Services* 71 (3): 243–249.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. "From Proof of Concept to Scalable Policies." *Journal of Economic Perspectives* 31 (4): 73–102.

Banerjee, Abhijit V., and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *The Annual Review of Economics* 1 (1):151–178.

Barden, Bret. 2013. *Assessing and Serving TANF Recipients with Disabilities*. OPRE Report 2013–56. Washington, DC: US Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

Barnow, Burt S. 1976. "The Use of Proxy Variables When One or Two Independent Variables Are Measured with Error." *American Statistician* 30 (3): 119–121.

Barnow, Burt S., and David Greenberg. 2015. "Do Estimated Impacts on Earnings Depend on the Source of the Data Used to Measure Them? Evidence from Previous Social Experiments." *Evaluation Review* 39 (2): 179–228.

Barnow, Burt S., and David Greenberg. 2019. "Special Issue Editors' Essay." *Evaluation Review* 43 (5): 231–265.

Barnow, Burt S., and David H. Greenberg. 2020. "Conducting Evaluations Using Multiple Trials." *American Evaluation Journal* 41 (4): 529–546.

Bell, Stephen H., and Laura R. Peck. 2016a. "On the Feasibility of Extending Social Experiments to Wider Applications." *Journal of MultiDisciplinary Evaluation* 12 (27): 93–112.

Bell, Stephen H., and Laura R. Peck. 2016b. "On the 'How' of Social Experiments: Experimental Designs for Getting Inside the Black Box." In *Social Experiments in Practice: The What, Why, When, Where, and How of Experimental Design & Analysis*, edited by Laura R. Peck, 97–109. Hoboken, NJ: Jossey-Bass.

Ben-Shalom, Yonatan, Steve Bruns, Kara Contreary, and David Stapleton. 2017. *Stay-at-Work/Return-to-Work: Key Facts, Critical Information Gaps, and Current Practices and Proposal*s. Washington, DC: Mathematica Policy Research.

Ben-Shalom, Yonatan, Jennifer Christian, and David Stapleton. 2018. "Reducing Job Loss among Workers with New Health Problems." In *Investing in America's Workforce: Improving Outcomes for Workers and Employers*, edited by Carl E. Van Horn, 267–288. Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Benítiz-Silva, Hugo, Moshe Buchinsky, and John Rust. 2010. "Induced Entry Effects of a \$1 for \$2 Offset in SSDI Benefits." Mimeo. https://editorialexpress.com/jrust/crest_lectures/induced_entry.pdf.

Berkowitz, E. D. 2013. *The Other Welfare: Supplemental Security Income and US Social Policy.* Ithaca, IL: Cornell University Press.

Berkowitz, Edward D. 2020. *Making Social Welfare Policy in America: Three Case Studies since 1950.* Chicago: University of Chicago Press.

Berkowitz, Edward D., and Larry DeWitt. 2013. *The Other Welfare: Supplemental Security Income and US Social Policy.* Ithaca, NY: Cornell University Press.

Bernanke, Ben. 2012. "The Federal Reserve and the Financial Crisis: Origins and Mission of the Federal Reserve, Lecture 1." Lecture presented at The George Washington University School of Business, Washington, DC, March 20. https://www.federalreserve.gov/mediacenter/files/chairman-bernanke-lecture1-20120320.pdf.

Bezanson, Birdie J. 2004. "The Application of Solution-Focused Work in Employment Counseling." *Journal of Employment Counseling* 41 (4): 183–191.

Biden, J. 2021. *Executive Order on Advancing Racial Equity and Support for Underserved Communities through the Federal Government.* EO 13985. Washington, DC: The White House.

Bitler, Marianne, P., Jonah B. Gelbach, and Hilary W. Hoynes. 2006. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." *American Economic Review* 96 (4): 988–1012.

Black, Dan, Kermit Daniel, and Seth Sanders. 2002. "The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust." *American Economic Review* 92 (1): 27–50.

Bloom, Howard S. 1984. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review* 8 (2): 225–246.

Bloom, Howard S. 1995. "Minimum Detectable Effects: A Simple Way to Report the Power of Experimental Designs." *Evaluation Review* 19 (5): 547–566.

Bloom, Howard S. 2009. *Modern Regression Discontinuity Analysis.* New York: MDRC.

Bloom, Howard S., Carolyn J. Hill, and James A. Riccio. 2003. "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments." *Journal of Policy Analysis and Management* 22 (4): 551–575.

Bloom, Howard S., Larry L. Orr, Stephen H. Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M. Bos. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study." *Journal of Human Resources* 32 (3): 549–576.

BLS (Bureau of Labor Statistics), US Department of Labor. 2019. "Characteristics of Unemployment Insurance Applicants and Benefit Recipients – 2018." News Release USDL-19-1692. https://www.bls.gov/news.release/pdf/uisup.pdf.

BLS (Bureau of Labor Statistics), US Department of Labor. 2020a. "Employee Access to Disability Insurance Plans." *The Economics Daily.* https://www.bls.gov/opub/ted/2018/employee-access-to-disability-insurance-plans.htm.

BLS (Bureau of Labor Statistics), US Department of Labor. 2020b. "Employer Reported Workplace Injuries and Illnesses – 2019." News Release USDL-20-2030. https://www.bls.gov/news.release/archives/osh_11042020.pdf.

Blustein, Jan. 2005. "Toward a More Public Discussion of the Ethics of Federal Social Program Evaluation." *Journal of Policy Analysis and Management* 24 (4): 824–846.

Board of Trustees, Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds. 2014. *The 2014 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds*. https://www.ssa.gov/OACT/TR/2014/.

Board of Trustees, Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds. 2019. *The 2019 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds*. Washington, DC: Author. https://www.ssa.gov/oact/tr/2019/tr2019.pdf.

Board of Trustees, Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds. 2021. *The 2021 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds.* Social Security Administration. https://www.ssa.gov/OACT/TR/2021/tr2021.pdf.

Boat, Thomas F., Stephen L. Buka, and James M. Perrin. 2015. "Children with Mental Disorders Who Receive Disability Benefits: A Report from the IOM." *Journal of the American Medical Association* 314 (19): 2019–2020.

Bond, Gary R. 1998. "Principles of the Individual Placement and Support Model: Empirical Support." *Psychiatric Rehabilitation Journal* 22 (1): 11–23.

Bond, G. R., D. R. Becker, and R. E. Drake. 2011. "Measurement of Fidelity of Implementation of Evidence-Based Practices: Case Example of the IPS Fidelity Scale." *Clinical Psychology: Science and Practice* 18: 126–141.

Bond, Gary R., Robert E. Drake, and Deborah R. Becker. 2008. "An Updated on Randomized Control Trials of Evidence-Based Supported Employment." *Psychiatric Rehabilitation. Journal* 31 (4): 280–290.

Bond, Gary R., Robert E. Drake, and Deborah R. Becker. 2012. "Generalizability of the Individual Placement and Support (IPS) Model of Supported Employment Outside the US." *World Psychiatry* 11 (1): 32--39.

Bond, Gary R., Robert E. Drake, Kim T. Mueser, and Eric Latimer. 2001. "Assertive Community Treatment for People with Severe Mental Illness." *Disease Management and Health Outcomes* 9 (3): 141–159.

Bond, Gary R., Robert E. Drake, and Jacqueline A. Pogue. 2019. "Expanding Individual Placement and Support to Populations 56hat Conditions and Disorders Other Than Serious Mental Illness." *Psychiatric Services* 70 (6): 488–498.

Bound, John. 1989. "The Health and Earnings of Rejected Disability Insurance Applicants." *American Economic Review* 79 (3): 482–503.

Bound, John. 1991. "The Health and Earnings of Disability Insurance Applicants: Reply." *American Economic Review* 81 (5): 1427–1434.

Bound, John, and Richard V. Burkhauser. 1999. "Economic Analysis of Transfer Programs Targeted on People with Disabilities." In *Handbook of Labor Economics*, vol. 3, edited by Orley Ashenfelter and David Card, 3417–3528. Amsterdam, The Netherlands: Elsevier.

Bound, John, Richard V. Burkhauser, and Austin Nichols. 2003. "Tracking the Household Income of SSDI and SSI Applicants." *Research in Labor Economics* 22: 113–158.

Bound, John, Julie Berry Cullen, Austin Nichols, and Lucie Schmidt. 2004. "The Welfare Implications of Increasing Disability Insurance Benefit Generosity." *Journal of Public Economics* 88 (12): 2487–2514.

Bound, John, Stephan Lindner, and Tim Waidmann. 2014. "Reconciling Findings on the Employment Effect of Disability Insurance." *IZA Journal of Labor Policy* 3 (1): 1–23.

Boyer, Sara L., and Gary R. Bond. 1999. "Does Assertive Community Treatment Reduce Burnout? A Comparison with Traditional Case Management." *Mental Health Services Research* 1 (1): 31–45.

Braitman, Alex, Peggy Counts, Richard Davenport, Barbara Zurlinden, Mark Rogers, Joe Clauss, Arun Kulkarni, Jerry Kymla, and Laura Montgomery. 1995. "Comparison of Barriers to Employment for Unemployed and Employed Clients in a Case Management Program: An Exploratory Study." *Psychiatric Rehabilitation Journal* 19 (1): 3–8.

Brock, Thomas, Michael J. Weiss, and Howard S. Bloom. 2013. *A Conceptual Framework for Studying the Sources of Variation in Program Effects.* New York: MDRC.

Brownson, Ross C., Amy A. Eyler, Jenine K. Harris, Justin B. Moore, and Rachel G. Tabak. 2018. "Getting the Word Out: New Approaches for Disseminating Public Health Science." *Journal of Public Health Management and Practice* 24 (2): 102–111.

Bruyere, Susanne M., Thomas P. Golden, and Ilene Zeitzer. 2007. "Evaluation and Future Prospect of U.S. Return to Work Policies for Social Security Beneficiaries." *Disability and Employment* 59: 53–90.

Burkhauser, Richard V., and Mary C. Daly. 2011. *The Declining Work and Welfare of People with Disabilities: What Went Wrong and a Strategy for Change.* Washington, DC: American Enterprise Institute Press.

Burkhauser, Richard V., Mary C. Daly, Duncan McVicar, and Roger Wilkins. 2014. "Disability Benefit Growth and Disability Reform in the US: Lessons from other OECD Nations." *IZA Journal of Labor Policy* 3 (4): 1–30.

Burstein, Nancy R., Cheryl A. Roberts, and Michelle L. Wood. 1999. *Recruiting SSA's Disability Beneficiaries for Return-to-Work: Results of the Project NetWork Demonstration: Final Report.* Bethesda, MD: Abt Associates.

Burtless, Gary. 1995. "The Case for Randomized Field Trials in Economic and Policy Research." *The Journal of Economic Perspectives* 9 (2): 63–84.

Burtless, Gary, and David Greenberg. 1982. "Inferences Concerning Labor Supply Behavior Based on Limited Duration Experiments." *The American Economic Review* 72 (3): 488–497.

Caliendo, Marco, and Sabine Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22 (1): 31–72.

Camacho, Christa Bucks, and Jeffrey Hemmeter. 2013. "Linking Youth Transition Support Services: Results from Two Demonstration Projects." *Social Security Bulletin* 73 (1). https://www.ssa.gov/policy/docs/ssb/v73n1/v73n1p59.html.

Campbell, Frances A., Elizabeth P. Pungello, Shari Miller-Johnson, Margaret Burchinal, and Craig T. Ramey. 2001. "The Development of Cognitive and Academic Abilities: Growth Curves from an Early Childhood Educational Experiment." *Developmental Psychology* 37 (2): 231–242.

Card, David, Jochen Kluve, and Andrea Weber. 2010. "Active Labour Market Policy Evaluations: A Meta-Analysis." *The Economic Journal* 120 (548): F452–F477.

Carter, Erik W., Diane Austin, and Audrey A. Trainor. 2012. "Predictors of Postschool Employment Outcomes for Young Adults with Severe Disabilities." *Journal of Disability Policy Studies* 23 (1): 50–63.

CBPP (Center on Budget and Policy Priorities). 2021. *Supplemental Security Income.* Policy Basics. Washington, DC: Author. https://www.cbpp.org/sites/default/files/atoms/files/PolicyBasics_SocSec-IntroToSSI.pdf.

CEA (Council of Economic Advisers). 2016. *Economic Report of the President, Transmitted to the Congress February 2016 Together with the Annual Report of the Council of Economic Advisors.* Washington DC: Government Printing Office.

CEP (Commission on Evidence-Based Policymaking). 2017. *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking.* Washington, DC: Author. https://bipartisanpolicy.org/wp-content/uploads/2019/03/Full-Report-The-Promise-of-Evidence-Based-Policymaking-Report-of-the-Comission-on-Evidence-based-Policymaking.pdf.

Chambless, Cathy, George Julnes, Sara McCormick, and Anne Brown-Reither. 2009. *Utah SSDI $1 for $2 Benefit Offset Pilot Demonstration Final Report.* Salt Lake City, UT: State of Utah.

Chambless, Catherine E., George Julnes, Sara T. McCormick, and Anne Reither. 2011. "Supporting Work Effort of SSDI Beneficiaries: Implementation of Benefit Offset Pilot Demonstration." *Journal of Disability Policy Studies* 22 (3): 179–188.

Charles, Kerwin Kofi, Yiming Li, and Melvin Stephens, Jr. 2018. "Disability Benefit Take-Up and Local Labor-Market Conditions." *Review of Economics and Statistics* 100 (3): 416–423.

Chetty, Raj. 2006. "A General Formula for the Optimal Level of Social Insurance." *Journal of Public Economics* 90 (10): 1879–1901.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The Fading American Dream: Trends in Absolute Income Mobility since 1940." *Science* 356 (6336): 398–406.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106 (4): 855–902.

Chow, Shein-Chung, and Mark Chang. 2012. *Adaptive Design Methods in Clinical Trials.* 2nd ed. Boca Raton, FL: CRC Press.

Christian, Jennifer, Thomas Wickizer, and A. Kim Burton. 2016. "A Community-Focused Health & Work Service (HWS)." In *SSDI Solutions: Ideas to Strengthen the Social Security Disability Insurance Program*, edited by Committee for a Responsible Federal Budget, The McCrery-Pomeroy SSDI Solutions Initiative, Ch. 4. Offprint. https://www.crfb.org/sites/default/files/christianwickizerburton.pdf.

Committee for a Responsible Federal Budget, The McCrery-Pomeroy SSDI Solutions Initiative. 2016. *SSDI Solutions: Ideas to Strengthen the Social Security Disability Insurance Program.* West Conshohocken, PA: Infinity Publishing.

Claes, Rita, and S. Antonio Ruiz-Quintanilla. 1998. "Influences of Early Career Experiences, Occupational Group, and National Culture on Proactive Career Behavior." *Journal of Vocational Behavior* 52 (3): 357–378.

Cloutier, Heidi, Joanne Malloy, David Hagner, and Patricia Cotton. 2006. "Choice and Control over Resources: New Hampshire's Individual Career Account Demonstration Projects." *Journal of Rehabilitation* 72 (2): 4–11.

Coldwell, Craig M., and William S. Bender. 2007. "The Effectiveness of Assertive Community Treatment for Homeless Populations with Severe Mental Illness: A Meta-Analysis." *American Journal of Psychiatry* 164 (3): 393–399.

Committee for the Prize in Economic Sciences in Memory of Alfred Nobel. 2019. *Understanding Development and Poverty Alleviation.* Stockholm, Sweden: The Royal Swedish Academy of Sciences.

Congressional Budget Office. 2012. *Policy Options for the Social Security Disability Insurance Program.* Washington, DC: Congress of the United States, Congressional Budget Office.

Cook, Thomas D. 2018. "Twenty-Six Assumptions That Have to Be Met If Single Random Assignment Experiments Are to Warrant 'Gold Standard' Status: A Commentary on Deaton and Cartwright." *Social Science & Medicine* 210: 37–40.

Cook, Thomas D., William R. Shadish, and Vivian C. Wong. 2008. "Three Conditions under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons." *Journal of Policy Analysis and Management* 27 (4): 724–750.

Cook, J., S. Shore, J. Burke-Miller, J. Jonikas, M. Hamilton, B. Ruckdeschel, et al. 2019. "Efficacy of Mental Health Self-Directed Care Financing in Improving Outcomes and Controlling Service Costs for Adults with Serious Mental Illness." *Psychiatric Services* 70 (3): 191–201.

Costa, Jackson. 2017. "The Decline in Earnings Prior to Application for Disability Insurance Benefits. *Social Security Bulletin* 77(1). https://www.ssa.gov/policy/docs/ssb/v77n1/v77n1p1.html.

Crepon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. "Do Labor Market Polices Have Displacement Effects? Evidence from a Clustered Randomized Experiment." *Quarterly Journal of Economics* 1238 (2): 531–580.

Cronbach, Lee J., Sueann Robinson Ambron, Sanford M. Dornbusch, Robert C. Hornik, D. C. Phillips, Decker F. Walker, and Stephen S. Winer. 1980. *Toward Reform of Program Evaluation.* San Francisco: Jossey-Bass.

Cunha, Flavio, and James J. Heckman. 2007. "The Evolution of Inequality, Heterogeneity, and Uncertainty in Labor Earnings in the US Economy." NBER Paper No. 13526. Cambridge, MA: National Bureau of Economic Research.

Cunha, Flavio, and James J. Heckman. 2008. "Formulating, Identifying, and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources* 43 (4): 738–782.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." NBER Paper No. 11331. Cambridge, MA: National Bureau of Economic Research.

Davies, Paul S., Kalman Rupp, and David Wittenburg. 2009. "A Life-Cycle Perspective on the Transition to Adulthood among Children Receiving Supplemental Security Income Payments." *Journal of Vocational Rehabilitation* 30 (3): 133–151.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine* 210: 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

Decker, Paul T., and Craig V. Thornton. 1995. "The Long-Term Effects of Transitional Employment Services." *Social Security Bulletin* 58 (4): 71–81.

Delin, Barry S., Ellie C. Hartman, and Christopher W. Sell. 2012. "The Impact of Work Outcomes: Evidence from Two Return-to-Work Demonstrations." *Journal of Vocational Rehabilitation* 36 (2): 97–107.

Delin, Barry S., Ellie C. Hartman, Christopher W. Sell, and Anne E. Brown-Reither. 2010. *Testing a SSDI Benefit Offset: An Evaluation of the Wisconsin SSDI Employment Pilot.* Menomonie, WI: University of Wisconsin-Stout.

Denne, Jacob, George Kettner, and Yonatan Ben-Shalom. 2015. *Return to Work in the Health Care Sector: Promising Practices and Success Stories.* Report for US Department of Labor, Office of Disability Employment Policy. Washington, DC: Mathematica Policy Research.

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

Deshpande, Manasi. 2016a. "Does Welfare Inhibit Success? I Long-Term Effects of Removing Low-Income Youth from the Disability Rolls." *American Economic Review* 106 (11): 3300–3330.

Deshpande, Manasi. 2016b. "The Effect of Disability Payments on Household Earnings and Income: Evidence from the SSI Children's Program." *Review of Economics and Statistics* 98 (4): 638–654.

Deshpande, Manasi. 2020. "How Disability Benefits in Early Life Affect Long-Term Outcomes." Center Paper NB20-05. Cambridge, MA: National Bureau of Economic Research.

Deshpande, Manasi, and Rebecca Dizon-Ross. 2020. *Improving the Outcomes of Disabled Youth through Information.* Cambridge, MA: National Bureau of Economic Research. https://grantome.com/grant/NIH/R21-HD091472-02.

DiClemente, Carlo C., James O. Prochaska, Scott K. Fairhurst, Wayne F. Velicer, Mary M. Velasquez, and Joseph S. Rossi. 1991. "The Process of Smoking Cessation: An Analysis of Precontemplation, Contemplation, and Preparation Stages of Change." *Journal of Consulting and Clinical Psychology* 59 (2): 295–304.

DiNardo, John, Jordan Matsudaira, Justin McCrary, and Lisa Sanbonmatsu. 2021. "A Practical Proactive Proposal for Dealing with Attrition: Alternative Approaches and an Empirical Example." *Journal of Labor Economics* 39 (S2): S507–S541.

Dixon, Lisa. 2000. "Assertive Community Treatment: Twenty-Five Years of Gold." *Psychiatric Services* 51 (6): 759–765.

Doemeland, Doerte, and James Trevino. 2014. "Which World Bank Reports Are Widely Read?" World Bank Policy Research Working Paper No. 6851. Washington, DC: The World Bank. http://documents1.worldbank.org/curated/en/387501468322733597/pdf/WPS6851.pdf.

DOL (US Department of Labor). 2015 [updated 2019]. *CLEAR Causal Evidence Guidelines, Version 2.1.* Washington, DC: US Department of Labor, Clearinghouse for Labor Evaluation and Research. https://clear.dol.gov/reference-documents/causal-evidence-guidelines-version-21.

DOL (US Department of Labor). n.d. "Employment First Presents 10 Critical Areas for Improving Competitive Integrated Employment Based on the WIOA Advisory Committee Report." Accessed December 10, 2020. https://www.dol.gov/sites/dolgov/files/odep/topics/employmentfirst/ef-presents-10-critical-areas-for-improving-cie-based-on-the-wioa-advisory-committee-report-full.pdf.

DOL (US Department of Labor). n.d. "RETAIN Initiative." Accessed September 24, 2021. https://www.dol.gov/agencies/odep/initiatives/saw-rtw/retain.

DOL (US Department of Labor). n.d. "WIOA Title I and III Annual Report Data: Program Year 2019." Workforce Performance Results, Employment and Training Administration. Accessed May 12, 2021. https://www.dol.gov/agencies/eta/performance/results.

DOL (US Department of Labor), ODEP (Office of Disability Employment Policy). 2018. "Notice of Availability of Funds and Funding Opportunity Announcement for: Retaining Employment and Talent after Injury/Illness Network Demonstration Projects." Issued May 24, 2018. https://www.dol.gov/sites/dolgov/files/odep/topics/saw-rtw/docs/foa-odep-18-01-published-on-grants.gov.pdf.

Dong, Nianbo, and Rebecca Maynard. 2013. "PowerUp! A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies." *Journal of Research on Educational Effectiveness* 6 (1): 24–67.

Duggan, Mark, and Scott A. Imberman. 2009. "Why Are the Disability Rolls Skyrocketing? The Contribution of Population Characteristics, Economic Conditions, and Program Generosity." In *Health at Older Ages*, edited by David M. Cutler and David A. Wise, 337–380. Chicago: University of Chicago Press.

Duggan, Mark G., and Melissa S. Kearney. 2007. "The Impact of Child SSI Enrollment on Household Outcomes." *Journal of Policy Analysis and Management* 26 (4): 861–885.

Duggan, Mark, Melissa S. Kearney, and Stephanie Rennane. 2015. "The Supplemental Income (SSI) Program." NBER Working Paper No. 21209. Cambridge, MA: National Bureau of Economic Research.

Duggan, Mark, Melissa S. Kearney, and Stephanie Rennane. 2016. "The Supplemental Security Income Program." In *Economics of Means-Tested Transfer Programs in the United Stated,* Vol. 2, edited by Robert A. Moffitt, 1–58. Chicago: University of Chicago Press.

Durlak, Joseph A., and Emily P. DuPre. 2008. "Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation." *American Journal of Community Psychology* 41 (3): 327-350.

Eeckhoudt, Louis, and Miles Kimball. 1992. "Background Risk, Prudence, and the Demand for Insurance." In *Contributions to Insurance Economics*, edited by Georges Dionne, 23–54. Boston: Kluwer Academic Publishers.

Eichengreen, Barry. 1996. *Golden Fetters: The Gold Standard and the Great Depression, 1919–1939*. New York: Oxford University Press.

Ekman, Lisa D. 2016. "Discussion of Early Intervention Proposals." In *SSDI Solutions: Ideas to Strengthen the Social Security Disability Insurance Program,* edited by Committee for a Responsible Federal Budget, The McCrery-Pomeroy SSDI Solutions Initiative, Ch. 3. Offprint. https://www.crfb.org/sites/default/files/stapletonbenshalommann.pdf.

Ellenhorn, Ross. 2005. "Parasuicidality and Patient Careerism: Treatment Recidivism and the Dialectics of Failure." *American Journal of Orthopsychiatry* 75 (2): 288–303.

Ellison, Marsha Langer, E. Sally Rogers, Ken Sciarappa, Mikal Cohen, and Rick Forbess. 1995. "Characteristics of Mental Health Case Management: Results of a National Survey." *The Journal of Mental Health Administration* 22 (2): 101–112.

Epstein, Diana, and Jacob Alex Klerman. 2012. "When Is a Program Ready for Rigorous Impact Evaluation? The Role of a Falsifiable Logic Model." *Evaluation Review* 36 (5): 375–401.

Epstein, Z., M. Wood, M. Grosz, S. Prenovitz, and A. Nichols. 2020. *Synthesis of Stay-at-Work/Return-to-Work (SAW/RTW) Programs, Models, Efforts, and Definitions.* Cambridge, MA: Abt Associates.

Farrell, Mary, Peter Baird, Bret Barden, Mike Fishman, and Rachel Pardoe. 2013. *The TANF/SSI Disability Transition Project: Innovative Strategies for Serving TANF Recipients with Disabilities*. OPRE Report 2013-51. Washington, DC: US Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

Farrell, Mary, and Johanna Walter. 2013. *The Intersection of Welfare and Disability: Early Findings from the TANF/SSI Disability Transition Project*. OPRE Report 2013-06. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, US Department of Health and Human Services.

Feely, Megan, Kristen D. Seay, Paul Lanier, Wendy Auslander, and Patricia L. Kohl. 2018. "Measuring Fidelity in Research Studies: A Field Guide to Developing a Comprehensive Fidelity Measurement System." *Child and Adolescent Social Work Journal* 35 (2): 139–152.

Fein, David, Samuel Dastrup, and Kimberly Burnett. 2021. *Still Bridging the Opportunity Divide for Low-Income Youth: Year Up's Longer-Term Impacts*. OPRE Report 2021-56. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, US Department of Health and Human Services. https://www.acf.hhs.gov/sites/default/files/documents/opre/year-up-report-april-2021.pdf.

Finkelstein, Amy, and Nathaniel Hendren. 2020. "Welfare Analysis Meets Causal Inference." *Journal of Economic Perspectives* 34 (4): 146–67. https://doi.org/10.1257/jep.34.4.146

Finkelstein, Amy, Sarah Taubman, Heidi Allen, Jonathan Gruber, Joseph P. Newhouse, Bill Wright, Kate Baicker, and Oregon Health Study Group. 2010. "*The Short-Run Impact of Extending Public Health Insurance to Low Income Adults: Evidence from the First Year of the Oregon Medicaid Experiment. Analysis Plan.* https://www.nber.org/sites/default/files/2020-02/analysis-plan-one-year-2010-12-01.pdf.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *The Quarterly Journal of Economics* 127 (3): 1057–1106.

Foster L., R. Brown, P. Phillips, J. Schore, and B. L. Carlson. 2003. "Improving the Quality of Medicaid Personal Assistance through Consumer Direction." *Health Affairs* 22 (Suppl 1). https://doi.org/10.1377/hlthaff.w3.162.

Foster, Jared C., Jeremy M. G. Taylor, and Stephen J. Ruberg. 2011. "Subgroup Identification from Randomized Clinical Trial Data." *Statistics in Medicine* 30 (24): 2867–2880. https://doi.org/10.1002/sim.4322.

Fraker, Thomas M., Peter Baird, Alison Black, Arif Mamun, Michelle Manno, John Martinez, Anu Rangarajan, and Debbie Reed. 2011. *The Social Security Administration's Youth Transition Demonstration Projects: Interim Report on Colorado Youth WIN.* Report for Social Security Administration, Office of Program Development and Research. Washington, DC: Mathematica Policy Research.

Fraker, Thomas, Peter Baird, Arif Mamun, John Martinez, Debbie Reed, and Allison Thompkins. 2012. *The Social Security Administration's Youth Transition Demonstration Projects: Interim Report on the Career Transition Program.* Center for Studying Disability Policy. Washington, DC: Mathematica Policy Research.

Fraker, Thomas, Alison Black, Joseph Broadus, Arif Mamun, Michelle Manno, John Martinez, Reanin McRoberts, Anu Rangarajan, and Debbie Read. 2011. *The Social Security Administration's Youth Transition Demonstration Projects: Interim Report on the City University of New York's Project.* Center for Studying Disability Policy. Washington, DC: Mathematica Policy Research.

Fraker, Thomas M., Alison Black, Arif Mamun, Michelle Manno, John Martinez, Bonnie O'Day, Meghan O'Toole, Anu Rangarajan, and Debbie Reed. 2011" *The Social Security Administration's Youth Transition Demonstration Projects: Interim Report on Transition WORK".* Report for Social Security Administration, Office of Program Development and Research. Washington, DC: Mathematica Policy Research.

Fraker, Thomas, Alison Black, Arif Mamun, John Martinez, Bonnie O'Day, Meghan O'Toole, Anu Rangarajan, and Debbie Read. 2011. *The Social Security Administration's Youth Transition Demonstration Projects: Interim Report on the Transition Works Project.* Center for Studying Disability Policy. Washington, DC: Mathematica Policy Research.

Fraker, Thomas, Erik Carter, Todd Honeycutt, Jacqueline Kauff, Gina Livermore, and Arif Mamun. 2014. *Promoting Readiness of Minors in SSI (PROMISE) Evaluation Design Report*. Washington, DC: Mathematica Policy Research.

Fraker, Thomas M., Joyanne Cobb, Jeffrey Hemmeter, Richard G. Luecking, and Arif Mamun. 2018. "Three-Year Effects of the Youth Transition Demonstration Projects." *Social Security Bulletin* 78 (3): 19–41.

Fraker, Thomas, Todd Honeycutt, Arif Mamun, Michelle Manno, John Martinez, Bonnie O'Day, Debbie Reed, and Allison Thompkins. 2012. *The Social Security Administration's Youth Transition Demonstration Projects: Interim Report on the Broadened Horizons, Brighter Futures.* Center for Studying Disability Policy. Washington, DC: Mathematica Policy Research.

Fraker, Thomas M., Richard G. Luecking, Arif A. Mamun, John M. Martinez, Deborah S. Reed, and David C. Wittenburg. 2016. "An Analysis of 1-Year Impacts of Youth Transition Demonstration Projects." *Career Development and Transition for Exceptional Individuals* 39 (1): 34–46.

Fraker, Thomas, Arif Mamun, Todd Honeycutt, Allison Thompkins, and Erin J. Valentine. 2014. *Final Report on the Youth Transition Demonstration*. Washington, DC: Mathematica Policy Research.

Fraker, Thomas, Arif Mamun, Michelle Manno, John Martinez, Debbie Reed, Allison Thompkins, and David Wittenburg. 2012. *The Social Security Administration's Youth Transition Demonstration Projects: Interim Report on the West Virginia Youth Works Project.* Center for Studying Disability Policy. Washington, DC: Mathematica Policy Research.

Fraker, Thomas, Arif Mamun, and Lori Timmins. 2015. *Three-Year Impacts of Services and Work Incentives on Youth with Disabilities*. Washington, DC: Mathematica Policy Research.

Fraker, Thomas, and Anu Rangarajan. 2009. "The Social Security Administration's Youth Transition Demonstration Projects." *Journal of Vocational Rehabilitation* 30 (3): 223–240.

Francesconi, Marco, and James J. Heckman. 2016. "Child Development and Parental Investment: Introduction." *The Economic Journal* 126 (596): F1–F27. https://doi.org/10.1111/ecoj.12388.

Frangakis, Constantine E., and Donald B. Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics* 58 (1): 21–29.

Franklin, Gary M., Thomas M. Wickizer, Norma B. Coe, and Deborah Fulton-Kehoe. 2015. "Workers' Compensation: Poor Quality Health Care and the Growing Disability Problem in the United States." *American Journal of Industrial Medicine* 58 (3): 245–251.

Freburger, Janet K., George M. Holmes, Robert P. Agans, Anne M. Jackman, Jane D. Darter, Andrea S. Wallace, Liana D. Castel, William D. Kalsbeek, and Timothy S. Carey. 2009. "The Rising Prevalence of Chronic Low Back Pain." *Archives of Internal Medicine* 169 (3): 251–258.

Freedman, Lily, Sam Elkin, and Megan Millenky. 2019. "Breaking Barriers: Implementing Individual Placement and Support in a Workforce Setting." New York: MDRC.

French, Eric, and Jae Song. 2014. "The Effect of Disability Insurance Receipt on Labor Supply." *American Economic Journal: Economic Policy* 6 (2): 291–337.

Frey, William D., Robert E. Drake, Gary R. Bond, Alexander L. Miller, Howard H. Goldman, David S. Salkever, Steven Holsenbeck, Mustafa Karakus, Roline Milfort, Jarnee Riley, Cheryl Reidy, Julie Bollmer, and Megan Collins. 2011. *Mental Health Treatment Study: Final Report*. Rockville, MD: Westat.

Fukui, Sadaaki, Rick Goscha, Charles A. Rapp, Ally Mabry, Paul Liddy, and Doug Marty. 2012. "Strengths Model Case Management Fidelity Scores and Client Outcomes." *Psychiatric Services* 63 (7): 708–710.

GAO (US Government Accountability Office). 2002. *Program Evaluation: Strategies for Assessing How Information Dissemination Contributes to Agency Goals*. Report No. GAO-02-923. Washington, DC: Author.

GAO (US Government Accountability Office). 2004. *Social Security Disability: Improved Processes for Planning and Conducting Demonstrations May Help SSA More Effectively Use Its Demonstration Authority*. Report No. GAO-05-19. Washington, DC: Author.

GAO (US Government Accountability Office). 2005. *Federal Disability Assistance, Wide Array of Programs Needs to Be Examined in Light of 21st Century Challenges*. Report No. GAO-05-626. Washington, DC: Author.

GAO (US Government Accountability Office). 2008. *Social Security Disability: Management Controls Needed to Strengthen Demonstration Projects*. Report No. GAO-08-1053. Washington, DC: Author.

GAO (US Government Accountability Office). 2010. *Highlights of a Forum: Actions That Could Increase Work Participation for Adults with Disabilities*. Report No. GAO-10-812SP. Washington, DC: Author.

GAO (US Government Accountability Office). 2012a. *Designing Evaluations: 2012 Revision*. Report No. GAO-12-208G. Washington, DC: Author.

GAO (US Government Accountability Office). 2012b. *Employment for People with Disabilities: Little Is Known about the Effectiveness of Fragmented and Overlapping Programs*. Report No. GAO-12-677. Washington, DC: Author.

GAO (US Government Accountability Office). 2012c. *Supplemental Security Income: Better Management Oversight Needed for Children's Benefits*. Report No. GAO-12-498SP. Washington, DC: Author.

GAO (US Government Accountability Office). 2017. *Supplemental Security Income: SSA Could Strengthen Its Efforts to Encourage Employment for Transition-Age Youth*. Report No. GAO-17-485. Washington, DC: Author.

GAO (US Government Accountability Office). 2018. *Medicaid Demonstrations: Evaluations Yielded Limited Results, Underscoring Need for Changes to Federal Policies and Procedures*. Report No. GAO-18-220. Washington, DC: Author.

GAO (US Government Accountability Office). 2019. *Medicaid Demonstrations: Approvals of Major Changes Need Increased Transparency*. Report No. GAO-19-315. Washington, DC: Author.

Gardiner, Karen N., and Randall Juras. 2019. *Pathways for Advancing Careers and Education: Cross-Program Implementation and Impact Study Findings*. OPRE Report 2019-32. Washington, DC: US Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

Gary, K. W., A. Sima, P. Wehman, and K. R. Johnson. 2019. "Transitioning Racial/Ethnic Minorities with Intellectual and Developmental Disabilities: Influence of Socioeconomic Status on Related Services." *Career Development and Transition for Exceptional Individuals* 42 (3): 158–167. https://doi.org/10.1177/2165143418778556.

Gelber, Alexander, Timothy J. Moore, and Alexander Strand. 2017. "The Effect of Disability Insurance Payments on Beneficiaries' Earnings." *American Economic Journal: Economic Policy* 9 (3): 229–261.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington, DC: The International Bank for Reconstruction and Development, The World Bank.

Geyer, Judy, Daniel Gubits, Stephen Bell, Tyler Morrill, Denise Hoffman, Sarah Croake, Katie Morrison, David Judkins, and David Stapleton. 2018. *BOND Implementation and Evaluation: 2017 Stage 2 Interim Process, Participation, and Impact Report.* Report for the Social Security Administration. Cambridge, MA: Abt Associates.

Gimm, Gilbert, Noelle Denny-Brown, Boyd Gilman, Henry T. Ireys, and Tara Anderson. 2009. *Interim Report on the National Evaluation of the Demonstration to Maintain Independence and Employment.* Washington, DC: Mathematica Policy Research.

Gingerich, Jade Ann, and Kelli Crane. 2021. *Transition Linkage Tool: A System Approach to Enhance Post-School Employment Outcomes.* Washington, DC: US Department of Labor, Office of Disability Employment Policy.

Gokhale, Jagadeesh. 2013" *A New Approach to SSDI Reform."* McCrery-Pomeroy SSDI Solutions Initiative Policy Brief. Washington, DC: Committee for a Responsible Federal Budget.

Gokhale, Jagadeesh. 2015. "SSDI Reform: Promoting Return to Work Without Compromising Economic Security." *Wharton Public Policy Initiative* 3 (7): 1–6.

Golden, Thomas P., Susan O'Mara, Connie Ferrell, and James R. Sheldon, Jr. 2000. "A Theoretical Construct for Benefits Planning and Assistance in the Ticket to Work and Work Incentive Improvement Act." *Journal of Vocational Rehabilitation* 14, (3): 147–152. https://content.iospress.com/articles/journal-of-vocational-rehabilitation/jvr00076.

Golden, T. P., S. O'Mara, C. Ferrell, J. Sheldon, and L. Axton Miller. 2005. *Supporting Career Development and Employment: Benefits Planning, Assistance and Outreach (BPA&O) and Protection and Advocacy for Beneficiaries of Social Security (PABSS).* SSA Publication No. 63-003. Social Security Administration. https://hdl.handle.net/1813/89921.

Goss, Steven C. 2013. *Testimony by Chief Actuary from Social Security Administration before the House Committee on Ways and Means, Subcommittee on Social Security*. Washington, DC: Social Security Administration.

Greenberg, David, Genevieve Knight, Stefan Speckesser, and Debra Hevenstone. 2011. "Improving DWP Assessment of the Relative Costs and Benefits of Employment Programmes." Working Paper No. 100. London, England: Department for Work and Pensions.

Greenberg, David, Robert H. Meyer, and Michael Wiseman. 1993. *Prying the Lid from the Black Box: Plotting Evaluation Strategy for Welfare Employment and Training Programs*. Madison, WI: University of Wisconsin-Madison, Institute for Research on Poverty.

Greenberg, David, Robert H. Meyer, and Michael Wiseman. 1994. "Multi-Site Employment and Training Evaluations: A Tale of Three Studies." *Industrial and Labor Relations Review* 47 (4): 679–691.

GSA (General Services Administration), OES (Office of Evaluation Sciences). 2018. *Increasing SSI Uptake: Letters to Adults 65 and Older Increased SSI Awards by 340%.* Washington, DC: Authors. https://oes.gsa.gov/assets/abstracts/1723-Increasing-SSI-Uptake.pdf.

GSA (General Services Administration), OES (Office of Evaluation Sciences). 2019a. *Communicating Employment Supports to Denied Disability Insurance Applicants.* https://oes.gsa.gov/assets/abstracts/15xx-di.pdf.

GSA (General Services Administration), OES (Office of Evaluation Sciences). 2019b. *Encouraging SSI Recipients to Self-Report Wage Changes.* Washington, DC: Authors. https://oes.gsa.gov/assets/abstracts/XXXX-ssi-wage-reporting-abstract.pdf.

GSA (General Services Administration), OES (Office of Evaluation Sciences). 2019c. "Encouraging SSI Recipients to Self-Report Wage Changes." https://oes.gsa.gov/projects/ssi-wage-reporting/.

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

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

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

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

Gubits, Daniel, Sarah Gibson, Michelle Wood, Cara Sierks, and Zachary Epstein. 2019. *Post-Entitlement Earnings Simplification Demonstration Technical Experts Panel Meeting: Final Report*. Rockville, MD: Abt Associates.

Guldi, Melanie, Amelia Hawkins, Jeffrey Hemmeter, and Lucie Schmidt. 2018. "Supplemental Security Income and Child Outcomes: Evidence from Birth Weight Eligibility Cutoffs." NBER Working Paper No. 24913. Cambridge, MA: National Bureau of Economic Research. https://www.nber.org/papers/w24913.

Hahn, Robert. 2019. "Building upon Foundations for Evidence-Based Policy," *Science* 364 (6440): 534–535.

Hall, Jean P., Catherine Ipsen, Noelle K. Kurth, Sara McCormick, and Catherine Chambless. 2020. "How Family Crises May Limit Engagement of Youth with Disabilities in Services to Support Successful Transitions to Postsecondary Education and Employment." *Children and Youth Services Review* 118: 1–7.

Hammermesh, Daniel S. 2007. "Viewpoint: Replication in Economics." *Canadian Journal of Economics* 40 (3): 715–733.

Heckman, James J. 1992. "Randomization and Social Policy Evaluation." In *Evaluating Welfare and Training Programs,* edited by Charles F. Manski and Irwin Garfinkel. Cambridge, MA: Harvard University Press.

Heckman, James J. 2011. "The Economics of Inequality: The Value of Early Childhood Education." *American Educator* 35, no. 1 (Spring): 31–47.

Heckman, James, Lance Lochner, and Ricardo Cossa. 2003. "Learning-by-Doing versus On-the-Job Training: Using Variation Induced by the EITC to Distinguish between Models of Skill Formation." In *Designing Social Inclusion: Tools to Raise Low-End Pay and Employment in Private Enterprise*, edited by Edmund S. Phelps, 74–130. Cambridge, United Kingdom: Cambridge University Press.

Heckman, James J., and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics* 6 (1): 689–733.

Heckman, James J., and Jeffrey A. Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9 (2): 85–110.

Heckman, James J., and Jeffrey A. Smith. 2004. "The Determinants of Participation in a Social Program: Evidence from a Prototypical Job Training Program." *Journal of Labor Economics* 22 (2): 243–298.

Heckman, James, Jeffrey Smith, and Christopher Taber. 1998. "Accounting for Dropouts in Evaluations of Social Programs." *The Review of Economics and Statistics* 80 (1): 1–14.

Heckman, J. J., and E. Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation 1." *Econometrica*, 73 (3): 669–738.

Hemmeter, Jeffrey. 2014. "Earnings and Disability Program Participation of Youth Transition Demonstration Participants after 24 Months." *Social Security Bulletin* 74 (1). https://www.ssa.gov/policy/docs/ssb/v74n1/v74n1p1.html.

Hemmeter, Jeffrey. 2015. "Supplemental Security Income Program Entry at Age 18 and Entrants' Subsequent Earnings." *Social Security Bulletin* 75 (3): 35–53.

Hemmeter, Jeffrey, and Michelle Stegman Bailey. 2016. "Earnings after DI: Evidence from Full Medical Continuing Disability Reviews." *IZA Journal of Labor Policy* 5 (1): 1–22.

Hemmeter, Jeffrey, and Joyanne Cobb. 2018. *Youth Transition Demonstration: Follow-Up Findings.* Presentation at the Fall Research Conference of the Association for Public Policy Analysis & Management, Washington, DC, November 2018.

Hemmeter, Jeffrey, Mark Donovan, Joyanne Cobb, and Tad Asbury. 2015. "Long Term Earnings and Disability Program Participation Outcomes of the Bridges Transition Program." *Journal of Vocational Rehabilitation* 42 (1): 1–15.

Hemmeter, Jeffrey, Michael Levere, Pragya Singh, and David Wittenburg. 2021. "Changing Stays? Duration of Supplemental Security Income Participation by First-Time Child Awardees and the Role of Continuing Disability Reviews." *Social Security Bulletin* 81 (2): 17–41.

Hemmeter, Jeffrey, David R. Mann, and David C. Wittenburg. 2017. "Supplemental Security Income and the Transition to Adulthood in the United States: State Variations in Outcomes Following the Age-18 Redetermination." *Social Service Review* 91 (1): 106–133.

Hemmeter, Jeffrey, John Phillips, Elana Safran, and Nicholas Wilson. 2020 "Communicating Program Eligibility: A Supplemental Security Income Field Experiment." Office of Evaluation Sciences Working Paper. https://oes.gsa.gov/assets/publications/1723%20-%20Hemmeter%20et%20al%20(2021)%20-%20Communicating%20Program%20Eligibility%20A%20Supplemental%20Security%20Income%20(SSI)%20Field%20Experiment.pdf.

Hemmeter, Jeffrey, and Michelle Stegman. 2015. "Childhood Continuing Disability Reviews and Age-18 Redeterminations for Supplemental Security Income Recipients: Outcomes and Subsequent Program Participation." *Research and Statistics Notes.* No. 2015-03. Social Security Administration. https://www.ssa.gov/policy/docs/rsnotes/rsn2015-03.html

Hendra, R., James A. Riccio, Richard Dorsett, David H. Greenberg, Genevieve Knight, Joan Phillips, Philip K. Robins, Sandra Vegeris, Johanna Walter, Aaron Hill, Kathryn Ray, and Jared Smith. 2011. *Breaking the Low-Pay, No-Pay Cycle: Final Evidence from the UK Employment Retention and Advancement (ERA) Demonstration.* Research Report No 765. London, England: Department for Work and Pensions.

Hendren, Nathaniel. 2016. "The Policy Elasticity." *Tax Policy and the Economy* 30 (1): 51–89.

Hendren, Nathaniel. 2020. "Measuring Economic Efficiency Using Inverse-Optimum Weights." NBER Working Paper No. 20351. Cambridge, MA: National Bureau of Economic Research. https://www.nber.org/papers/w20351.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2019. "Unified Welfare Analysis of Government Policies." NBER WP No. 26144. https://www.nber.org/papers/w26144.

Herd, Pamela, and Donald P. Moynihan. 2018. *Administrative Burden: Policymaking by Other Means*. New York: Russell Sage Foundation.

Hernandez, Brigida, Mary J. Cometa, Jay Rosen, Jessica Velcoff, Daniel Schober, and Rene D. Luna. 2006. "Employment, Vocational Rehabilitation, and the Ticket to Work Program: Perspectives of Latinos with Disabilities." *Journal of Applied Rehabilitation Counseling* 37 (3): 13–22.

HHS/ACF/OPRE (US Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation). 2020. *Portfolio of Research in Welfare and Family Self-Sufficiency*. OPRE Report 2021-13. Washington, DC: Author.

Higgins, Julian P.T., and Simon G. Thompson. 2004. "Controlling the Risk of Spurious Findings from Meta-Regression." *Statistics in Medicine* 23 (11): 1663–1682.

Hill, Fiona. 2020. "Public Service and the Federal Government." *Policy 2020 Voter Vitals.* Washington, DC: Brookings Institution.

Hirano, Kara A., Dawn Rowe, Lauren Lindstrom, and Paula Chan. 2018. "Systemic Barriers to Family Involvement in Transition Planning for Youth with Disabilities: A Qualitative Metasynthesis." *Journal of Child and Family Studies* 27 (11): 3440–3456.

Hock, Heinrich, Michael Levere, Kenneth Fortson, and David Wittenburg. 2019. *Lessons from Pilot Tests of Recruitment for the Promoting Opportunity Demonstration.* Report for Social Security Administration, Office of Research, Demonstration, and Employment Support. Washington, DC: Mathematica Policy Research.

Hock, Heinrich, Dara Lee Luca, Tim Kautz, and David Stapleton. 2017. *Improving the Outcomes of Youth with Medical Limitations through Comprehensive Training and Employment Services: Evidence from the National Job Corps Study.* Washington, DC: Mathematica Policy Research.

Hock, Heinrich, David Wittenburg, and Michael Levere. 2020. "Memorandum: Promoting Opportunity Demonstration: Recruitment and Random Assignment Report." Washington, DC: Mathematica Policy Research.

Hock, Heinrich, David Wittenburg, Michael Levere, Noelle Denny-Brown, and Heather Gordon. 2020. *Promoting Opportunity Demonstration: Recruitment and Random Assignment Report.* Washington, DC: Mathematica Policy Research.

Hoffman, Denise, Sarah Croake, David R. Mann, David Stapleton, Priyanka Anand, Chris Jones, Judy Geyer, Daniel Gubits, Stephen Bell, Andrew McGuirk, David Wittenburg, Debra Wright, Amang Sukasih, David Judkins, and Michael Sinclair. 2017. *2016 Stage 1 Interim Process, Participation, and Impact Report.* Report for the Social Security Administration (contract deliverable 24c2.1 under Contract SS00-10-60011), Office of Program Development & Research. Cambridge, MA: Abt Associates; and Washington, DC: Mathematica Policy Research.

Hoffman, Denise, Jeffrey Hemmeter, and Michelle S. Bailey. 2018. "The Relationship between Youth Services and Adult Outcomes among Former Child SSI Recipients." *Journal of Vocational Rehabilitation* 48 (2): 233–247.

Hoffmann, Holger, Dorothea Jäckel, Sybille Glauser, Kim T. Mueser, and Zeno Kupper. 2014. "Long-Term Effectiveness of Supported Employment: 5-Year Follow-Up of a Randomized Controlled Trial." *American Journal of Psychiatry* 171 (11): 1183–1190.

Holbrook, Allyson L., Timothy P. Johnson, and Maria Krysan. 2019. "Race- and Ethnicity-of-Interviewer Effects." In *Experimental Methods in Survey Research: Techniques That Combine Random Sampling with Random Assignment*, edited by Paul Lavrakas, Michael Traugott, Courtney Kennedy, Allyson Holbrook, Edith de Leeuw, and Brady West, 197–224. Hoboken, NJ: John Wiley & Sons.

Hollenbeck, Kevin. 2015. *Promoting Retention or Reemployment of Workers after a Significant Injury or Illness.* Report for US Department of Labor, Office of Disability Employment Policy. Washington, DC: Mathematica Policy Research.

Hollenbeck, K. 2021. *Demonstration Evidence of Early Intervention Policies and Practices.* Kalamazoo, MI: W. E. Upjohn Institute.

Hollister, Robinson G., Peter Kemper, and Rebecca A Maynard. 1984. *The National Supported Work Demonstration.* Madison, WI: University of Wisconsin Press.

Holt, Stephen, and Katie Vinopal. 2021. "It's About Time: Examining Inequality in the Time Cost of Waiting." *SSRN.* https://doi.org/10.2139/ssrn.3857883.

Honeycutt, Todd, Kara Contreary, and Gina Livermore. 2021. *Considerations for the Papers Developed for the SSI Youth Solutions Project.* Report for the US Department of Labor, Office of Disability Employment Policy. Princeton, NJ: Mathematica. https://www.mathematica.org/publications/considerations-for-the-papers-developed-for-the-ssi-youth-solutions-project.

Honeycutt, Todd, Brittney Gionfriddo, Jacqueline Kauff, Joseph Mastrianni, Nicholas Redel, and Adele Rizzuto. 2018. *Promoting Readiness of Minors in Supplemental Security Income (PROMISE): Arkansas PROMISE Process Analysis Report.* Washington, DC: Mathematica Policy Research.

Honeycutt, Todd, Brittney Gionfriddo, and Gina Livermore. 2018. *Promoting Readiness of Minors in Supplemental Security Income (PROMISE): PROMISE Programs' Use of Effective Transition Practices in Serving Youth with Disabilities.* Washington, DC: Mathematica Policy Research.

Honeycutt, Todd, and Gina Livermore. 2018. *Promoting Readiness in Minors in Supplemental Security Income (PROMISE): The Role of PROMISE in the Landscape of Federal Programs Targeting Youth with Disabilities.* Washington, DC: Mathematica Policy Research.

Honeycutt, Todd, Eric Morris, and Thomas Fraker. 2014. *Preliminary YTD Benefit-Cost Analysis Using Administrative Data.* Princeton, NJ: Mathematica Policy Research.

Honeycutt, T., and Stapleton, D. 2013. "Striking While the Iron Is Hot: The Effect of Vocational Rehabilitation Service Wait Times on Employment Outcomes for Applicants Receiving Social Security Disability Benefits." *Journal of Vocational Rehabilitation* 39 (2): 137–152.

Honeycutt, Todd, David Wittenburg, Kelli Crane, Michael Levere, Richard Luecking, and David Stapleton. 2018. *Supplemental Security Income Youth Formative Research Project: Considerations for Identifying Promising and Testable Interventions.* Washington, DC: Mathematica Policy Research.

Honeycutt, Todd, David Wittenburg, Michael Levere, and Sarah Palmer. 2018. *Supplemental Security Income Youth Formative Research Project: Target Population Profiles.* Washington, DC: Mathematica Policy Research.

Hotz, V. Joseph, and John Karl Scholz. 2001. "Measuring Employment Income for Low-Income Populations with Administrative and Survey Data." In *Studies of Welfare Populations: Data Collection and Research Issues,* edited by M. V. Ploeg, R. A. Moffit, and C. F. Citro, 275–315. Washington, DC: The National Academies Press.

Hotz, V. J., and J. K. Scholz. 2003. "The Earned Income Tax Credit." In *Means-Tested Transfer Programs in the United States,* edited by R. Moffitt, 141–198. Chicago: University of Chicago Press.

Hoynes, H. W., and R. Moffitt. 1999. "Tax Rates and Work Incentives in the Social Security Disability Insurance Program: Current Law and Alternative Reforms." *National Tax Journal* 52 (4): 623–654.

Huggett, Mark, Gustavo Ventura, and Amir Yaron. 2011. "Sources of Lifetime Inequality." *American Economic Review* 101 (7): 2923–2954.

Hullegie, Patrick, and Pierre Koning. 2015. "Employee Health and Employer Incentives." Discussion Paper No. 9310. Bonn, Germany: Institute for the Study of Labor.

Hussey, Michael A., and James P. Hughes. 2007. "Design and Analysis of Stepped Wedge Cluster Randomized Trials." *Contemporary Clinical Trials* 28 (2): 182–191.

IAIABC (International Association of Industrial Accident Boards and Commissions), Disability Management and Return to Work Committee. 2016. *Return to Work: A Foundational Approach to Return to Function.* Madison, WI: Author.

Ibarraran, Pablo, Laura Ripani, Bibiana Taboada, Juan Miguel Villa, and Brigida Garcia. 2014. "Life Skills, Employability, and Training for Disadvantaged Youth: Evidence from a Randomized Evaluation Design." *IZA Journal of Labor & Development* 3 (1): 1–24.

Imai, K., D. Tingley, and T. Yamamoto. 2013. "Experimental Designs for Identifying Causal Mechanisms." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176 (1): 5-51.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001.

Imbens, Guido W., and Donald B. Rubin. 2015. *An Introduction to Causal Inference in Statistics, Biomedical and Social Sciences.* New York: Cambridge University Press.

Inanc, Hande, and David R. Mann. 2019. "Recent Changes and Reforms to the United Kingdom's Income Support Program for People with Disabilities." Center for Studying Disability Policy, Working Paper 2019-16. Washington, DC: Mathematica.

Iwanaga, Kanako, Paul Wehman, Valerie Brooke, Lauren Avellone, and Joshua Taylor. 2021. "Evaluating the Effect of Work Incentives Benefits Counseling on Employment Outcomes of Transition-Age and Young Adult Supplemental Security Income Recipients with Intellectual Disabilities: A Case Control Study." *Journal of Occupational Rehabilitation* 31: 581-591.

Johnson, George E. 1979. "The Labor Market Displacement Effect in the Analysis of the Net Impact of Manpower Training Programs." *Research in Labor Economics*, Supplement 1, 227–254.

Johnson, George E., and James D. Tomola. 1977. "The Fiscal Substitution Effect of Alternative Approaches to Public Service Employment Policy." *Journal of Human Resources* 12 (1): 3–26.

Kanter, Joel. 1989. "Clinical Case Management: Definition, Principles, Components." *Psychiatric Services* 40 (4): 361–368.

Kapteyn, Arie, and Jelmer Y. Ypma. 2007. "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data." *Journal of Labor Economics* 25 (3): 513–551.

Karhan, Andrew J., and Thomas P. Golden. 2021. *Policy Considerations for Implementing Youth and Family Case Management Strategies across Systems.* Washington, DC: US Department of Labor, Office of Disability Employment Policy.

Katz, Lawrence F. 1994. "Active Labor Market Policies to Expand Employment and Opportunity." In *Reducing Unemployment: Current Issues and Policy Options*, 239–290. Jackson Hole, WY: Federal Reserve Bank of Kansas City.

Kauff, Jacqueline, Jonathan Brown, Norma Altshuler, and Noelle Denny-Brown. 2009. *Findings from a Study of the SSI/SSDI Outreach, Access, and Recovery (SOAR) Initiative.* Washington, DC: Mathematica Policy Research.

Kauff, Jacqueline F., Elizabeth Clary, Kristin Sue Lupfer, and Pamela J. Fischer. 2016. "An Evaluation of SOAR: Implementation and Outcomes of an Effort to Improve Access to SSI and SSDI." *Psychiatric Services* 67 (10): 1098–1102.

Kauff, Jacqueline, Elizabeth Clary, and Julia Lyskawa. 2014. *An Evaluation of SOAR: The Implementation and Outcomes of an Effort to Increase Access to SSI and SSDI.* Washington, DC: Mathematica Policy Research.

Kauff, Jacqueline, Todd Honeycutt, Karen Katz, Joseph Mastrianni, and Adele Rizzuto. 2018. *Promoting Readiness of Minors in Supplemental Security Income (PROMISE): Maryland PROMISE Process Analysis Report.* Washington, DC: Mathematica Policy Research.

Kennedy, Courtney, and Hannah Hartig. 2019. "Response Rates in Telephone Surveys Have Resumed Their Decline" (blog), *Pew Research Center.* February 27, 2019. https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/.

Kennedy, Elizabeth, and Laura King. 2014. "Improving Access to Benefits for Persons with Disabilities Who Were Experiencing Homelessness: An Evaluation of the Benefits Entitlement Services Team Demonstration Project." *Social Security Bulletin* 74 (4): 45–55.

Kerachsky, Stuart, and Craig Thornton. 1987. "Findings from the STETS Transitional Employment Demonstration." *Exceptional Children* 53 (6): 515–521.

Kerachsky, Stuart, Craig Thornton, Anne Bloomenthal, Rebecca Maynard, and Susan Stephens. 1985. *Impacts of Transitional Employment on Mentally Retarded Young Adults: Results of the STETS Demonstration.* Washington, DC: Mathematica Policy Research.

Kerksick, Julie, David Riemer, and Conor Williams. 2016. "Using Transitional Jobs to Increase Employment of SSDI Applicants and Beneficiaries." In *SSDI Solutions: Ideas to Strengthen the Social Security Disability Insurance Program*, edited by Committee for a Responsible Federal Budget, The McCrery-Pomeroy SSDI Solutions Initiative, Ch. 5. West Conshohocken, PA: Infinity Publishing.

Kimball, Miles S. 1990. "Precautionary Saving in the Small and in the Large." *Econometrica* 58 (1): 53–73.

King, Gary, and Richard Nielsen. 2019. "Why Propensity Scores Should Not Be Used for Matching" *Political Analysis* 27 (4): 435–454.

Klerman, Jacob. 2020. "Findings from the (Experimental) Job Training Literature." Abt Associates. Mimeo.

Kluve, Jochen, Susana Puerto, David Robalino, Jose Maunel Romero, Friederike Rother, Jonathan Stöterau, Felix Weidenkaff, and Marc Witte. 2016. "Do Youth Employment Programs Improve Labor Market Outcomes? A Systematic Review." IZA Discussion Paper, No. 10263. Bonn, Germany: Institute for the Study of Labor. https://ftp.iza.org/dp10263.pdf.

Knaus, Michael C., Michael Lechner, and Anthony Strittmatter. 2020. "Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach." *Journal of Human Resources*. https://doi.org/10.3368/jhr.57.2.0718-9615R1.

Ko, Hansoo, Renata E. Howland, and Sherry A. Glied. 2020. "The Effects of Income on Children's Health: Evidence from Supplemental Security Income Eligibility under New York State Medicaid." NBER Working Paper No. 26639. Cambridge, MA: National Bureau of Economic Research. https://www.nber.org/papers/w26639.

Kogan, Deborah, Hannah Betesh, Marian Negoita, Jeffrey Salzman, Laura Paulen, Haydee Cuza, Liz Potamites, Jillian Berk, Carrie Wolfson, and Patty Cloud. 2012. *Evaluation of the Senior Community Service Employment Program (SCSEP) Process and Outcomes Study Final Report.* Report for US Department of Labor, Employment and Training Administration, Office of Policy Development and Research. Oakland, CA: Social Policy Research Associates.

Kornfeld, Robert, and Kalman Rupp. 2000. "The Net Effects of the Project NetWork Return-to-Work Case Management Experiment on Participant Earnings, Benefit Receipt, and Other Outcomes." *Social Security Bulletin* 63 (1): 12–33.

Kornfeld, Robert J., Michelle L. Wood, Larry L. Orr, and David A. Long. 1999. *Impacts of the Project NetWork Demonstration: Final Report.* Report for Social Security Administration. Bethesda, MD: Abt Associates.

Kregel, John. 2006a. *Conclusions Drawn from the State Partnership Initiative.* Richmond, VA: Virginia Commonwealth University, Rehabilitation Research and Training Center, State Partnership Systems Change Initiative Project Office. https://www.ssa.gov/disabilityresearch/documents/spiconclusions.pdf.

Kregel, John. 2006b. *Final Evaluation Report of the SSI Work Incentives Demonstration Project.* Richmond, VA: Virginia Commonwealth University, Rehabilitation Research and Training Center, State Partnership Systems Change Initiative Project Office. https://www.ssa.gov/disabilityresearch/documents/spireport.pdf.

Kregel, John, and Susan O'Mara. 2011. "Work Incentive Counseling as a Workplace Support." *Journal of Vocational Rehabilitation* 35 (2): 73–83. https://www.doi.org/10.3233/JVR-2011-0555.

Kunz, Tanja, and Marek Fuchs. 2019. "Using Experiments to Assess Interactive Feedback That Improves Response Quality in Web Surveys." In Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment, edited by Paul Lavrakas, Michael Traugott, Courtney Kennedy, Allyson Holbrook, Edith de Leeuw, and Brady West, 247–274. Hoboken, NJ: John Wiley & Sons.

Larson, Sheryl A., and Judy Geyer. 2021. "Delaying Application of SSI's Substantial Gainful Activity Eligibility Criterion from Age 18 to 22." Washington, DC: US Department of Labor, Office of Disability Employment Policy.

Lavrakas, Paul J., Jenny Kelly, and Colleen McClain. 2019. "Investigating Interviewer Effects and Confounds in Survey-Based Experimentation." In *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment,* edited by Paul Lavrakas, Michael Traugott, Courtney Kennedy, Allyson Holbrook, Edith de Leeuw, and Brady West, 225–244. Hoboken, NJ: John Wiley & Sons.

Leiter, Valerie, Michelle L. Wood, and Stephen H. Bell. 1997. "Case Managements at Work for SSA Disability Beneficiaries: Process Results of the Project NetWork Return-to-Work Demonstration." *Social Security Bulletin* 60: 29–48.

Levere, Michael, Todd Honeycutt, Gina Livermore, Arif Mamun, and Karen Katz. 2020. *Family Service Use and Its Relationship with Youth Outcomes.* Washington, DC: Mathematica Policy Research.

Levy, Frank. 1979. "The Labor Supply of Female Household Heads, or AFDC Work Incentives Don't Work Too Well." *Journal of Human Resources* 14 (1): 76–97.

Liebman, Jeffrey B. 2015. "Understanding the Increase in Disability Insurance Benefit Receipt in the United States." *Journal of Economic Perspectives* 29 (2): 123–150.

Liebman, Jeffrey B., and Jack A. Smalligan. 2013. "Proposal 4: An Evidence-Based Path to Disability Insurance Reform." In *15 Ways to Rethink the Federal Budget*, 27–30. Washington, DC: The Hamilton Project.

Liu, Su, and David C. Stapleton. 2011. "Longitudinal Statistics on Work Activity and Use of Employment Supports for New Social Security Disability Insurance Beneficiaries." *Social Security Bulletin* 71 (3): 35–59.

Livermore, Gina. 2011. "Social Security Disability Beneficiaries with Work-Related Goals and Expectations." *Social Security Bulletin* 71 (3): 61–82.

Livermore, Gina A., and Nanette Goodman. 2009. *A Review of Recent Evaluation Efforts Associated with Programs and Policies Designed to Promote the Employment of Adults with Disabilities.* Princeton, NJ: Mathematica Policy Research.

Livermore, Gina, Todd Honeycutt, Arif Mamun, and Jacqueline Kauff. 2020. "Insights about the Transition System for SSI Youth from the National Evaluation of Promoting Readiness of Minors in SSI (PROMISE)." *Journal of Vocational Rehabilitation* 52 (1): 1–17.

Livermore, Gina, Arif Mamun, Jody Schimmel, and Sarah Prenovitz. 2013. *Executive Summary of the Seventh Ticket to Work Evaluation Report.* Washington, DC: Mathematica Policy Research.

Livermore, Gina, and Sarah Prenovitz. 2010. *Benefits Planning, Assistance, and Outreach (BPAO) Service User Characteristics and Use of Work Incentives*. Work Activity and Use of Employment Supports under the Original Ticket to Work Regulations, Final Report. No. 5ca13079097b4ae887f19a614aca2bec. Washington, DC: Mathematica Policy Research.

Livermore, Gina, David Wittenburg, and David Neumark. 2014. "Finding Alternatives to Disability Benefit Receipt." *IZA Journal of Labor Policy* 3 (14). https://doi.org/10.1186/2193-9004-3-14.

Lowenstein, Amy E., Noemi Altman, Patricia M. Chou, Kristen Faucetta, Adam Greeney, Daniel Gubits, Jorgen Harris, JoAnn Hsueh, Erika Lundquist, Charles Michalopoulos, and Vinh Q. Nguyen. 2014. *A Family-Strengthening Program for Low-Income Families: Final Impacts from the Supporting Healthy Marriage Evaluation, Technical Supplement*. OPRE Report 2014-09B. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services.

Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan. 2011. "Mechanism Experiments and Policy Evaluations." *Journal of Economic Perspectives* 25 (3): 17–38.

Luecking, Richard G., and David C. Wittenburg. 2009. "Providing Supports to Youth with Disabilities Transitioning to Adulthood: Case Descriptions from the Youth Transition Demonstration." *Journal of Vocational Rehabilitation*, 30: 241–251.

Maestas, Nicole. 2019. "Identifying Work Capacity and Promoting Work: A Strategy for Modernizing the SSDI Program." *The ANNALS of the American Academy of Political and Social Science* 686 (1): 93–120.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review* 103 (5): 1797–1829.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. Forthcoming. "The Effect of Economic Conditions on the Disability Insurance Program: Evidence from the Great Recession." *Journal of Public Economics.*

Maestas, Nicole, Kathleen J. Mullen, and Gema Zamarro. 2010. *Research Designs for Estimating Induced Entry into the SSDI Program Resulting from a Benefit Offset.* Santa Monica, CA: The RAND Corporation.

Malani, Anup. 2006. "Identifying Placebo Effects with Data from Clinical Trials." *Journal of Political Economy* 114 (2): 236–256.

Mamun, Arif, Ankita Patnaik, Michael Levere, Gina Livermore, Todd Honeycutt, Jacqueline Kauff, Karen Katz, AnnaMaria McCutcheon, Joseph Mastrianni, and Brittney Gionfriddo. 2019. *Promoting Readiness of Minors in SSI (PROMISE) Evaluation: Interim Services and Impact Report*. Washington, DC: Mathematica Policy Research.

Mamun, Arif, David Wittenburg, Noelle Denny-Brown, Michael Levere, David R. Mann, Rebecca Coughlin, Sarah Croake, Heather Gordon, Denise Hoffman, Rachel Holzwat, Rosalind Keith, Brittany McGill, and Aleksandra Wec. 2021. *Promoting Opportunity Demonstration: Interim Evaluation Report.* Report for Social Security Administration, Office of Research, Demonstration, and Employment Support. Washington, DC: Mathematica Policy Research.

Manchester, Joyce. 2019. *Targeting Early Intervention Based on Health Care Utilization of SSDI Beneficiaries by State, with Emphasis on Mental Disorders and Substance Abuse.* Washington, DC: Committee for a Responsible Federal Budget, McCrery-Pomeroy SSDI Solutions Initiative. https://www.crfb.org/sites/default/files/Targeting_Early_Intervention_Based_On_Health_Care_Utilization.pdf.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2013. "Poverty Impedes Cognitive Function." *Science* 341 (6149): 976–980.

Marrow Jocelyn, Daley Tamara, Taylor Jeffrey, Karakus Mustafa, Marshall Tina, Lewis Megan. 2020. *Supported Employment Demonstration. Interim Process Analysis Report (Deliverable 7.5a).* Rockville, MD: Westat. https://www.ssa.gov/disabilityresearch/documents/SED_Interim_Process_Analysis_Report_8-07-20.pdf.

Martin, F., and Sevak, P. 2020. "Implementation and Impacts of the Substantial Gainful Activity Project Demonstration in Kentucky." *Journal of Vocational Rehabilitation* (Preprint), 1-9.

Martin, Patricia P. 2016. "Why Researchers Now Rely on Surveys for Race Data on OASDI and SSI Programs: A Comparison of Four Major Surveys." *Research and Statistics Notes.* No. 2016-01. Social Security Administration. https://www.ssa.gov/policy/docs/rsnotes/rsn2016-01.html.

Martinez, John, Thomas Fraker, Michelle Manno, Peter Baird, Arif Mamun, Bonnie O'Day, Anu Rangarajan, David Wittenburg, and Social Security Administration. 2010. *Social Security Administration's Youth Transition Demonstration Projects: Implementation Lessons from the Original Sites*. Washington, DC: Mathematica Policy Research.

Martinson, Karin, Doug McDonald, Amy Berninger, and Kyla Wasserman. 2021. *Building Evidence-Based Strategies to Improve Employment Outcomes for Individuals with Substance Use Disorders.* OPRE Report 2020-171. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, US Department of Health and Human Services.

Matulewicz, Holly, Karen Katz, Todd Honeycutt, Jacqueline Kauff, Joseph Mastrianni, Adele Rizzuto, and Claire S. Wulsin. 2018. *Promoting Readiness of Minors in Supplemental Security Income (PROMISE): California PROMISE Process Analysis Report*. Washington, DC: Mathematica Policy Research.

Maximus. 2002. *Youth Continuing Disability Review Project: Annual Report October 1, 2001–September 30, 2002*. Report to the Social Security Administration, Office of Employment Support Programs.

McCann, Ted, and Nick Hart. 2019. "Disability Policy: Saving Disability Insurance with the First Reforms in a Generation." In *Evidence Works: Cases Where Evidence Meaningfully Informed Policy*, edited by Nick Hart and Meron Yohannes, 28–39. Washington, DC: Bipartisan Policy Center.

McConnell, Sheena, and Steven Glazerman. 2001. *National Job Corps Study: The Benefits and Costs of Job Corps*. Washington, DC: Mathematica Policy Research.

McConnell, Sheena, Irma Perez-Johnson, and Jillian Berk. 2014. "Proposal 9: Providing Disadvantaged Workers with Skills to Succeed in the Labor Market." In *Policies to Address Poverty in America*, edited by Melissa S. Kearney and Benjamin H. Harris, 97–189. Washington, DC: The Brookings Institution.

McCoy, Marion L., Cynthia S. Robins, James Bethel, Carina Tornow, and William D. Frey. 2007. *Evaluation of Homeless Outreach Projects and Evaluation: Task 6: Final Evaluation Report*. Rockville, MD: Westat.

McCutcheon, AnnaMaria, Karen Katz, Rebekah Selekman, Todd Honeycutt, Jacqueline Kauff, Joseph Mastrianni, and Adele Rizzuto. 2018. *Promoting Readiness of Minors in Supplemental Security Income (PROMISE): New York State PROMISE Process Analysis Report*. Washington, DC: Mathematica Policy Research.

McHugo, G. J., R. E. Drake, R. Whitley, G. R. Bond, K. Campbell, C. A. Rapp, H. H. Goldman, W. J. Lutz, and M. T. Finnerty. 2007. "Fidelity Outcomes in the National Implementing Evidence-Based Practices Project." *Psychiatric Services* 58: 1279–1284.

McLaughlin, James R. 1994. "Estimated Increase in OASDI Benefit Payments That Would Result from Two 'Earnings Test' Type Alternatives to the Current Criteria for Cessation of Disability Benefits—Information." Memorandum, SSA Office of the Actuary.

Metcalf, C. E. 1973. "Making Inferences from Controlled Income Maintenance Experiments." *American Economic Review* 63 (3): 478–483.

Meyer, Bruce D. 1995. "Lessons from the US Unemployment Insurance Experiments." *Journal of Economic Literature* 33 (1): 91–131.

Meyers, Marcia K., Janet C. Gornick, and Laura R. Peck. 2002. "More, Less, or More of the Same? Trends in State Social Welfare Policy in the 1990s." *Publius: The Journal of Federalism* 32 (4): 91–108.

Michalopoulos, Charles, David Wittenburg, Dina A. R. Israel, Jennifer Schore, Anne Warren, Aparajita Zutshi, Stephen Freedman, and Lisa Schwartz. 2011. *The Accelerated Benefits Demonstration and Evaluation Project: Impacts on Health and Employment at Twelve Months*. New York: MDRC. http://www.ssa.gov/disabilityresearch/documents/AB%20Vol%201_508%20comply.pdf.

Miller, L., and S. O'Mara. 2003 [updated 2004]. "Social Security Disability Benefit Issues Affecting Transition Aged Youth." Briefing Paper, vol. 8. Richmond, VA: Virginia Commonwealth University, Benefits Assistance Resource Center.

Moffitt, Robert A. 1992a. "Evaluation Methods for Program Entry Effects." In *Evaluating Welfare and Training Programs*, edited by C. F. Manski and I. Garfinkel, 231–252. Cambridge, MA: Harvard University Press.

Moffitt, Robert. 1992b. "Incentive Effects of the US Welfare System: A Review." *Journal of Economic Literature* 30 (1): 1–61.

Moffitt, Robert A. 1996. "The Effect of Employment and Training Programs on Entry and Exit from the Welfare Caseload." *Journal of Policy Analysis and Management* 15 (1): 32–50.

Moffitt, Robert, ed. 2016. *Economics of Means-Tested Transfer Programs in the United States*. Chicago: University of Chicago Press.

Mojtabai, Ramin. 2011. "National Trends in Mental Health Disability, 1997–2009." *American Journal of Public Health* 101 (11): 2156–2163.

Moynihan, Donald, Pamela Herd, and Hope Harvey. 2015. "Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions." *Journal of Public Administration Research and Theory* 25 (1): 43–69.

Mullen, Kathleen J., and Stephanie L. Rennane. 2017. "The Effect of Unconditional Cash Transfers on the Return to Work of Permanently Disabled Workers." NBER Working Paper No. DRC NB17-09. Cambridge, MA: National Bureau of Economic Research: https://www.nber.org/programs-projects/projects-and-centers/retirement-and-disability-research-center/center-papers/drc-nb17-09.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2015. *Mental Disorders and Disabilities among Low-Income Children.* Washington, DC: The National Academies Press. https://doi.org/10.17226/21780.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2018. *Opportunities for Improving Programs and Services for Children with Disabilities.* Washington, DC: The National Academies Press.

National Association of Social Work. 2013. "NASW Standards for Social Work Case Management." https://www.socialworkers.org/LinkClick.aspx?fileticket=acrzqmEfhlo%3D&portalid=0.

National Disability Institute. 2020. *Race, Ethnicity, and Disability: The Financial Impact of Systemic Inequality and Intersectionality.* Washington, DC: Author. https://www.nationaldisabilityinstitute.org/wp-content/uploads/2020/08/race-ethnicity-and-disability-financial-impact.pdf.

National Safety Council. 2020. "NSC Injury Facts." https://injuryfacts.nsc.org/.

Nazarov, Zafar. 2013. "Can Benefits and Work Incentives Counseling Be a Path to Future Economic Self-Sufficiency for SSI/SSDI Beneficiaries?" Working Paper No. 2013-17. Chestnut Hill, MA: Center for Retirement Research at Boston College.

NCWD/Y (National Collaborative on Workforce and Disability for Youth). 2005. *Guideposts for Success.* Washington, DC: Institute on Education Leadership, 2005.

NCWD/Y (National Collaborative on Workforce and Disability for Youth). 2009. *Guideposts for Success,* 2nd ed. Washington, DC: Institute on Educational Leadership.

NCWD/Y (National Collaborative on Workforce and Disability for Youth). 2019. *Guideposts for Success 2.0: A Framework for Successful Youth Transition to Adulthood.* Washington, DC: Author. http://www.ncwd-youth.info/wp-content/uploads/2019/07/Guideposts-for-Success-2.0.pdf.

Neuhauser, Frank. 2016, April. "The Myth of Workplace Injuries: Or Why We Should Eliminate Workers' Compensation for 90% of Workers and Employers." *IAIABC Perspectives*. https://resources.iaiabc.org/1a4arng/.

Nichols, Austin, Emily Dastrup, Zachary Epstein, and Michelle Wood. 2020. *Data Analysis for Stay-at-Work/Return-to-Work (SAW/RTW) Models and Strategies Project. Early Intervention Pathway Map and Population Profiles.* Report for US Department of Labor. Cambridge, MA: Abt Associates.

Nichols, A., J. Geyer, M. Grosz, Z. Epstein, and M. Wood. 2020. *Synthesis of Evidence about Stay-at-Work/ Return-to-Work (SAW/RTW) and Related Programs.* Report for the U.S. Department of Labor. Rockville, MD: Abt Associates.

Nichols, Austin, and Jesse Rothstein. 2016. "The Earned Income Tax Credit." In *Economics of Means-Tested Transfer Programs in the United States*, Vol. 1, edited by Robert A. Moffitt, 137–218. Chicago: University of Chicago Press.

Nichols, Austin, Lucie Schmidt, and Purvi Sevak. 2017. "Economic Conditions and Supplemental Security Income Applications." *Social Security Bulletin* 77 (4): 27–44.

Nickow, Andre, Philip Oreopoulos, and Vincent Quan. 2020. "The Impressive Effects of Tutoring on Prek–12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence." NBER Working Paper No. 27476. Cambridge, MA: National Bureau of Economic Research.

Noel, Valerie A., Eugene Oulvey, Robert E. Drake, Gary R. Bond, Elizabeth A. Carpenter-Song, and Brian DeAtley. 2018. "A Preliminary Evaluation of Individual Placement and Support for Youth with Developmental and Psychiatric Disabilities." *Journal of Vocational Rehabilitation* 48 (2): 249–255.

NTACT (National Technical Assistance Center on Transition). 2016. *Evidence-Based Practices and Predictors in Secondary Transition: What We Know and What We Still Need to Know*. Charlotte, NC: Author. https://transitionta.org/wp-content/uploads/docs/EBPP_Exec_Summary_2016_12-13.pdf.

Nunn, Ryan, Jana Parsons, and Jay Shambaugh. 2019. *Labor Force Nonparticipation: Trends, Causes, and Policy Solutions*. The Hamilton Project. Washington, DC: Brookings. https://www.hamiltonproject.org/assets/files/PP_LFPR_final.pdf.

Nye-Lengerman, Kelly, Amy Gunty, David Johnson, and Maureen Hawes. 2019. "What Matters: Lessons Learned from the Implementation of PROMISE Model Demonstration Projects." *Journal of Vocational Rehabilitation* 51 (2): 275–284.

O'Day, Bonnie, Hannah Burak, Kathleen Feeney, Elizabeth Kelley, Frank Martin, Gina Freeman, Grace Lim, and Katie Morrison. 2016. *Employment Experiences of Young Adults and High Earners Who Receive Social Security Disability Benefits: Findings from Semistructured Interviews*. Washington, DC: Mathematica Policy Research.

O'Day, Bonnie, Allison Roche, Norma Altshuler, Liz Clary, and Krista Harrison. 2009. *Process Evaluation of the Work Incentives Planning and Assistance Program.* Work Activity and Use of Employment Supports under the Original Ticket to Work Regulations, Report 1. Washington, DC: Mathematica Policy Research.

O'Leary, Paul, Leslie I. Boden, Seth A. Seabury, Al Ozonoff, and Ethan Scherer. 2012. "Workplace Injuries and the Take-Up of Social Security Disability Benefits." *Social Security Bulletin* 72 (3): 1–17.

Olney, Marjorie F., and Cindy Lyle. 2011. "The Benefits Trap: Barriers to Employment Experienced by SSA Beneficiaries." *Rehabilitation Counseling Bulletin* 54 (4): 197–209.

Olsen, Anya, and Russell Hudson. 2009. "Social Security Administration's Master Earnings File: Background Information," *Social Security Bulletin* 69 (3): 29–46.

Olsen, Robert B., Larry L. Orr, Stephen H. Bell, and Elizabeth A. Stuart. 2013. "External Validity in Policy Evaluations That Choose Sites Purposively." *Journal of Policy Analysis and Management* 32 (1): 107–121. https://doi.org/10.1002/pam.21660.

Orr, Larry L. 1999. *Social Experiments: Evaluating Public Programs with Experimental Methods*. Thousand Oaks, CA: Sage.

Page, Lindsay C., Avi Feller, Todd Grindal, Luke Miratrix, and Marie-Andree Somers. 2015. "Principal Stratification: A Tool for Understanding Variation in Program Effects across Endogenous Subgroups." *American Journal of Evaluation* 36 (4): 514–531.

Parsons, Donald O. 1980. "The Decline in Male Labor Force Participation." *Journal of Political Economy* 88 (1): 117–134.

Peck, Laura R. 2003. "Subgroup Analysis in Social Experiments: Measuring Program Impacts Based on Post Treatment Choice." *American Journal of Evaluation* 24 (2): 157–187.

Peck, Laura R. 2005. "Using Cluster Analysis in Program Evaluation." *Evaluation Review* 29: (25): 178–196.

Peck, Laura R. 2013. "On Analysis of Symmetrically Predicted Endogenous Subgroups: Part One of a Method Note in Three Parts." *American Journal of Evaluation* 34 (2): 225–236.

Peck, Laura R. 2020. *Experimental Evaluation Design for Program Improvement*. Thousand Oaks, CA: Sage.

Peck, Laura R., Daniel Litwok, Douglas Walton, Eleanor Harvill, and Alan Werner. 2019. *Health Profession Opportunity Grants (HPOG 1.0) Impact Study: Three-Year Impacts Report*. OPRE Report 2019-114. Report for US Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. Rockville, MD: Abt Associates.

Peck, Laura R., and Ronald J. Scott, Jr. 2005. "Can Welfare Case Management Increase Employment? Evidence from a Pilot Program Evaluation." *Policy Studies Journal* 33 (4): 509–533.

Peikes, Deborah N., Lorenzo Moreno, and Sean Michael Orzol. 2008. "Propensity Score Matching: A Note of Caution for Evaluators of Social Programs." *The American Statistician* 62 (3): 222–231.

Peikes, Deborah, Sean Orzol, Lorenzo Moreno, and Nora Paxton. 2005. *State Partnership Initiative: Selection of Comparison Groups for the Evaluation and Selected Impact Estimates: Final Report*. Princeton, NJ: Mathematica Policy Research.

The Policy Surveillance Program. n.d. "State Supplemental Payments for Children with Disabilities." Accessed September 20, 2021. http://www.lawatlas.org/datasets/supplemental-security-income-for-children-with-disabilities.

Porter, Alice, James Smith, Alydia Payette, Tim Tremblay, and Peter Burt. 2009. *SSDI $1 for $1 Benefit Offset Pilot Demonstration Vermont Pilot Final Report*. Burlington, VT: Vermont Division of Vocational Rehabilitation. https://www.ssa.gov/disabilityresearch/documents/Vt1for2FinalReport091223.pdf.

Prero, Aaron J., and Craig Thornton. 1991. "Transitional Employment Training for SSI Recipients with Mental Retardation." *Social Security Bulletin* 54 (11): 2–25.

Proudlock, S., and N. Wellman. 2011. "Solution Focused Groups: The Results Look Promising." *Counselling Psychology Review* 26 (3): 45–54.

Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price. 2009. "What to Do When Data Are Missing in Group Randomized Controlled Trials." NCEE 2009-0049. Washington, DC: US Department of Education.

Rangarajan, Anu, Thomas Fraker, Todd Honeycutt, Arif Mamun, John Martinez, Bonnie O'Day, and David Wittenburg. 2009. *The Social Security Administration's Youth Transition Demonstration Projects: Evaluation Design Report*. No. dc181046c9a041e6b63bb1b5743e1935. Princeton, NJ: Mathematica Policy Research.

Rothstein, Jesse, and Till von Wachter. 2017. "Social Experiments in the Labor Market." In *Handbook of Economic Field Experiments*, Vol. 2, edited by Abhijit Vinayak Banerjee and Esther Duflo, 555–637. Amsterdam, The Netherlands: North-Holland/Elsevier.

Ruiz-Quintanilla, S. Antonio, Robert R. Weathers II, Valerie Melburg, Kimberly Campbell, and Nawaf Madi. 2006. "Participation in Programs Designed to Improve Employment Outcomes for Persons with Psychiatric Disabilities: Evidence from the New York WORKS Demonstration Project." *Social Security Bulletin* 66 (2): 49–79.

Rupp, Kalman, Stephen H. Bell, and Leo A. McManus. 1994. "Design of the Project NetWork Return-to-Work Experiment for Persons with Disabilities." *Social Security Bulletin* 57: 3. (2): 3–20. https://pubmed.ncbi.nlm.nih.gov/7974091/.

Rupp, Kalman, Michelle Wood, and Stephen H. Bell. 1996. "Targeting People with Severe Disabilities for Return-to-Work: The Project NetWork Demonstration Experience." *Journal of Vocational Rehabilitation* 7 (1–2): 63–91.

SAMHSA (Substance Abuse and Mental Health Services Administration). n.d. "SSI/SSDI Outreach, Access and Recovery: An Overview." Rockville, MD: Author. https://soarworks.samhsa.gov/sites/soarworks.prainc.com/files/SOAROverview-2020-508_0.pdf.

Sampson, James P., Robert C. Reardon, Gary W. Peterson, and Janet G. Lenz. 2004. *Career Counseling and Services: A Cognitive Information Processing Approach*. Belmont, CA: Thomson/Brooks/Cole.

Schiller, Bradley R. 1973. "Empirical Studies of Welfare Dependency: A Survey." *Journal of Human Resources* 8: 19–32.

Schimmel, Jody, David Stapleton, David Mann, and Dawn Phelps. 2013. *Participant and Provider Outcomes since the Inception of Ticket to Work and the Effects of the 2008 Regulatory Changes*. Report for Social Security Administration, Office of Research, Demonstration, and Employment Support. Washington, DC: Mathematica Policy Research.

Schimmel, Jody, David C. Stapleton, and Jae G. Song. 2011. "How Common Is Parking among Social Security Disability Insurance Beneficiaries. Evidence from the 1999 Change in the Earnings Level of Substantial Gainful Activity." *Social Security Bulletin* 71 (4): 77–92.

Schlegelmilch, Amanda, Matthew Roskowski, Cayte Anderson, Ellie Hartman, and Heidi Decker-Maurer. 2019. "The Impact of Work Incentives Benefits Counseling on Employment Outcomes of Transition-Age Youth Receiving Supplemental Security Income (SSI) Benefits." *Journal of Vocational Rehabilitation* 51 (2): 127–136.

Schmidt, Lucie, and Purvi Sevak. 2004. "AFDC, SSI, and Welfare Reform Aggressiveness." *Journal of Human Resources* 39 (3): 792–812.

Schmidt, Lucie, and Purvi Sevak. 2017. "Child Participation in Supplemental Security Income: Cross- and within-State Determinants of Caseload Growth." *Journal of Disability Policy Studies* 28 (3): 131–140.

Schmidt, Lucie, Lara D. Shore-Sheppard, and Tara Watson. 2020. "The Impact of the ACA Medicaid Expansion on Disability Program Applications." *American Journal of Health Economics* 6 (4): 444–476.

Schochet, Peter Z. 2009. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Evaluation Review* 33 (6): 539–567.

Schochet, Peter Z., John Burghardt, and Sheena McConnell. 2006. *National Job Corps Study and Longer-Term Follow-Up Study: Impact and Benefit-Cost Findings Using Survey and Summary Earnings Records Data. Final Report*. Princeton, NJ: Mathematica Policy Research.

Schochet, Peter Z., Sheena M. McConnell, and John A. Burghardt. 2003. *National Job Corps Study: Findings Using Administrative Earnings Records Data*. Princeton, NJ: Mathematica Policy Research, Inc.

Selekman, Rebekah, Mary A. Anderson, Todd Honeycutt, Karen Katz, Jacqueline Kauff, Joseph Mastrianni, and Adele Rizzuto. 2018. *Promoting Readiness of Minors in Supplemental Security Income (PROMISE): Wisconsin PROMISE Process Analysis Report*. Washington, DC: Mathematica Policy Research.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth/Cengage Learning.

Skidmore, Sara, Debra Wright, Kirsten Barrett, and Eric Grau. 2017. *National Beneficiary Survey–General Waves Round 5*. Vol. 2: *Data Cleaning and Identification of Data Problems*. Washington, DC: Mathematica.

Smalligan, Jack, and Chantel Boyens. 2019. "Improving the Social Security Disability Determination Process." Washington, DC: Urban Institute.

Smalligan, Jack, and Chantel Boyens. 2020. "Two Proposals to Strengthen Paid-Leave Programs." Washington, DC: Urban Institute.

Smith, Jeffrey A., and Petra E. Todd. 2005. "Does Matching Overcome LaLonde's Critique of Non-Experimental Estimators?" *Journal of Econometrics* 125 (1–2): 305–353.

Social Security Advisory Board. 2016. "Representative Payees: A Call to Action." *Issue Brief.* https://www.ssab.gov/research/representative-payees-a-call-to-action/.

Solomon, Phyllis. 1992. "The Efficacy of Case Management Services for Severely Mentally Disabled Clients." *Community Mental Health Journal* 28 (3): 163–180.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. "What Are We Weighting For?" *Journal of Human Resources* 50 (2): 301–316.

SRI International. 1983. *Final Report of the Seattle-Denver Income Maintenance Experiment.* Vol. 1, *Design and Results.* Washington, DC: Government Printing Office.

SSA (Social Security Administration). 2001. "Childhood Disability: Supplemental Security Income Program. A Guide for Physicians and Other Health Care Professionals." Social Security Administration. https://www.ssa.gov/disability/professionals/childhoodssi-pub048.htm.

SSA (Social Security Administration). 2006. "Cooperative Agreements for Work Incentives Planning and Assistance Projects; Program Announcement No. SSA-OESP-06-1." *Federal Register.* https://www.federalregister.gov/documents/2006/05/16/06-4507/program-cooperative-agreements-for-work-incentives-planning-and-assistance-projects-program.

SSA (Social Security Administration). 2016. *The Social Security Administration's Plan to Achieve Self-Support Program.* Audit Report A-08-16-50030. Office of the Inspector General. https://oig-files.ssa.gov/audits/full/A-08-16-50030.pdf.

SSA (Social Security Administration). 2018a. *National Beneficiary Survey: Disability Statistics, 2015.* Baltimore, MD: Author.

SSA (Social Security Administration). 2018b. *Social Security Programs throughout the World: Europe, 2018.* SSA Publication No. 13-11801. Washington, DC: Social Security Administration, Office of Research, Evaluation, and Statistics, Office of Retirement and Disability Policy.

SSA (Social Security Administration). 2019a. *Annual Report on Medical Continuing Reviews: Fiscal Year 2015.* Baltimore, MD: Author. https://www.ssa.gov/legislation/FY%202015%20CDR%20Report.pdf.

SSA (Social Security Administration). 2019b. *Annual Report on Section 234 Demonstration Projects.* Washington, DC: Author. https://www.ssa.gov/disabilityresearch/documents/Section%20234%20Report%20-%202019.pdf.

SSA (Social Security Administration). 2019c. *Annual Statistical Report on the Social Security Disability Insurance Program, 2018.* Washington, DC: Author. https://www.ssa.gov/policy/docs/statcomps/di_asr/2018/di_asr18.pdf.

SSA (Social Security Administration). 2019d. "Supplemental Security Income, Table 7.B1." Annual Statistical Supplement. http://www.ssa.gov/policy/docs/statcomps/supplement/2019/7b.html#table7.b1.

SSA (Social Security Administration). 2020a. *Annual Report on Section 234 Demonstration Projects*. Baltimore, MD: Author. https://www.ssa.gov/legislation/Demo%20Project%20Report%20Released%20-%20Section%20234%20Report%202020.pdf.

SSA (Social Security Administration). 2020b. *Annual Statistical Report on the Social Security Disability Insurance Program, 2019.* https://www.ssa.gov/policy/docs/statcomps/di_asr/2019/di_asr19.pdf.

SSA (Social Security Administration). 2020c. *Annual Statistical Supplement to the Social Security Bulletin*. Baltimore, MD: Author.

SSA (Social Security Administration). 2020d. *DI & SSI Program Participants: Characteristics & Employment, 2015*. Washington, DC: Author. https://www.ssa.gov/policy/docs/chartbooks/di-ssi-employment/2015/dsppce-2015.pdf.

SSA (Social Security Administration). 2020e. *Red Book. A Summary Guide to Employment Supports for People with Disabilities under the Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) Programs.* https://www.ssa.gov/redbook/.

SSA (Social Security Administration). 2020f, September. *Social Security Administration Evaluation Policy.* Washington, DC: Author. https://www.ssa.gov/data/data_governance_board/Evidence%20Act%20Evaluation%20Policy%20-%20September%202020.pdf.

SSA (Social Security Administration). 2020g. *SSA Budget Information.* https://www.ssa.gov/budget/FY21Files/2021BO.pdf.

SSA (Social Security Administration). 2020h. *SSI Annual Statistical Report, 2019*. Washington, DC: Author. https://www.ssa.gov/policy/docs/statcomps/ssi_asr/2019/ssi_asr19.pdf.

SSA (Social Security Administration). 2020i. *What You Need to Know about Your Supplemental Security Income (SSI) When You Turn 18*. Report No. 2020. Baltimore, MD: Author. www.socialsecurity.gov/pubs/EN-05-11005.pdf.

SSA (Social Security Administration). 2021. "SSI Monthly Statistics, 2020." Research, Statistics & Policy Analysis. https://www.ssa.gov/policy/docs/statcomps/ssi_monthly/2020/index.html.

SSA (Social Security Administration). n.d. "Requesting an Electronic Data Exchange with SSA." Accessed March 26, 2021. https://www.ssa.gov/dataexchange/request_dx.html.

SSA (Social Security Administration). n.d. "State Vocational Rehabilitation Agency Reimbursements." VR Reimbursement Claims Processing website. https://www.ssa.gov/work/claimsprocessing.html (accessed May 7, 2021).

SSA (Social Security Administration). n.d. "Ticket Tracker, August 2020." Accessed March 4, 2021. https://www.ssa.gov/work/tickettracker.html.

SSA/ORDP/ORDES (Social Security Administration; Office of Retirement and Disability Policy; Office of Research, Demonstration, and Employment Support). 2020. *Overview and Documentation of the Social Security Administration's Disability Analysis File (DAF) Public Use File for 2019.* Washington, DC: Mathematica. Retrieved from https://www.ssa.gov/disabilityresearch/daf_puf.html#documentation.

Stapleton, David C., Stephen H. Bell, Denise Hoffman, and Michelle Wood. 2020. "Comparison of Population-Representative and Volunteer Experiments: Lessons from the Social Security Administration's Benefit Offset National Demonstration (BOND)." *American Journal of Evaluation* 41 (4): 547–563.

Stapleton, David, Stephen Bell, David Wittenburg, Brian Sokol, and Debi McInnis. 2010. *BOND Implementation and Evaluation: BOND Final Design Report.* Report for Social Security Administration. Washington, DC: Abt Associates.

Stapleton, David, Yonatan Ben-Shalom, and David Mann. 2016. "The Employment/Eligibility System: A New Gateway for Employment Supports and Social Security Disability Benefits." In *SSDI Solutions: Ideas to Strengthen the Social Security Disability Insurance Program*, edited by Committee for a Responsible Federal Budget, The McCrery-Pomeroy SSDI Solutions Initiative, Ch. 3. Offprint. https://www.crfb.org/sites/default/files/ stapletonbenshalommann.pdf.

Stapleton, David, Yonatan Ben-Shalom, and David R. Mann. 2019. *Development of an Employment/Eligibility Services (EES) System.* Report for University of New Hampshire. Washington, DC: Mathematica Policy Research.

Stapleton, David, Robert Burns, Benjamin Doornink, Mary Harris, Robert Anfield, Winthrop Cashdollar, Brian Gifford, and Kevin Ufier. 2015. *Targeting Early Intervention to Workers Who Need Help to Stay in the Labor Force.* Report for US Department of Labor, Office of Disability Employment Policy. Washington, DC: Mathematica Policy Research.

Stapleton, David, Arif Mamun, and Jeremy Page. 2014. "Initial Impacts of the Ticket to Work Program: Estimates Based on Exogenous Variation in Ticket Mail Months." *IZA Journal of Labor Policy* 3 (1): 1–24.

State of Connecticut. 2009. *Benefit Offset Pilot Demonstration: Connecticut Final Report.* Report for Social Security Administration. https://www.ssa.gov/ disabilityresearch/documents/Conn-FINAL%20BOP%20REPORT%2012%207%2009.doc.

Stepner, Michael. 2019. "The Long-Term Externalities of Short-Term Disability Insurance." Unpublished working paper. https://files.michaelstepner.com/ short_term_di_externalities.pdf.

Stuart, Elizabeth A., Stephen R. Cole, Catherine P. Bradshaw, and Philip J. Leaf. 2011. "The Use of Propensity Scores to Assess the Generalizability of Results from Randomized Trials." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 174 (2): 369–386.

Taylor, Jeffrey, David Salkever, William Frey, Jarnee Riley, and Jocelyn Marrow. 2020. *Supported Employment Demonstration Final Enrollment Analysis Report (Deliverable 7.4b).* Report for Social Security Administration. Rockville, MD: Westat.

Test, David W., Valerie L. Mazzotti, April L. Mustian, Catherine H. Fowler, Larry Kortering, and Paula Kohler. 2009. "Evidence-Based Secondary Transition Predictors for Improving Postschool Outcomes for Students with Disabilities." *Career Development for Exceptional Individuals* 32 (3): 160–181.

Thornton, Craig, and Paul Decker. 1989. *The Transitional Employment Training Demonstration: Analysis of Program Impacts.* Princeton, NJ: Mathematica Policy Research.

Thornton, Craig, Shari Miller Dunstan, and Jennifer Schore. 1988. *The Transitional Employment and Training Demonstration: Analysis of Program Operations.* Princeton, NJ: Mathematica Policy Research.

Thornton, Craig, Gina Livermore, Thomas Fraker, David Stapleton, Bonnie O'Day, David Wittenburg, Robert Weathers II, et al. 2007. *Evaluation of the Ticket to Work Program: Assessment of Post-Rollout Implementation and Early Impacts,* Vol. 1. Washington, DC: Mathematica Policy Research.

Tipton, Elizabeth. 2013. "Improving Generalizations from Experiments Using Propensity Score Subclassification: Assumptions, Properties, and Contexts" *Journal of Educational and Behavioral Statistics* 38 (3): 239–266.

Tipton, Elizabeth. 2014. "How Generalizable Is Your Experiment? An Index for Comparing Experimental Samples and Populations." *Journal of Educational and Behavioral Statistics* 39 (6): 478–501.

Tipton, Elizabeth, and Laura R. Peck. 2017. "A Design-Based Approach to Improve External Validity in Welfare Policy Evaluations." *Evaluation Review* 41 (4): 326–356.

Tipton, Elizabeth, David S. Yeager, Ronaldo Iachan, and Barbara Schneider. 2019. "Designing Probability Samples to Study Treatment Effect Heterogeneity." In *Experimental Methods in Survey Research: Techniques That Combine Random Sampling with Random Assignment,* edited by Paul Lavrakas, Michael Traugott, Courtney Kennedy, Allyson Holbrook, Edith de Leeuw, and Brady West, 435–456. Hoboken, NJ: John Wiley & Sons.

Todd, Petra E., and Kenneth I. Wolpin. 2006. "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility." *American Economic Review* 96 (5): 1384–1417.

Tremblay, Tim, James Smith, Alice Porter, and Robert Weathers. 2011. "Effects on Beneficiary Employment and Earnings of a Graduated $1-for-$2 Benefit Offset for Social Security Disability Insurance (SSDI)." *Journal of Rehabilitation* 77 (2): 19.

Tremblay, T., J. Smith, H. Xie, and R. Drake. 2004. "The Impact of Specialized Benefits Counseling Services on Social Security Administration Disability Beneficiaries in Vermont." *Journal of Rehabilitation* 70 (2): 5-11.

Tremblay, Timothy, James Smith, Haiyi Xie, and Robert E. Drake. 2006. "Effect of Benefits Counseling Services on Employment Outcomes for People with Psychiatric Disabilities." *Psychiatric Services* 57 (6): 816–821.

Trepper, Terry S., Yvonne Dolan, Eric E. McCollum, and Thorana Nelson. 2006. "Steve De Shazer and the Future of Solution-Focused Therapy." *Journal of Marital and Family Therapy* 32 (2): 133–139.

Treskon, Louisa. 2016. "What Works for Disconnected Young People: A Scan of the Evidence." MDRC Working Paper. New York: MDRC.

Tuma, Nancy B. 2001. "Approaches to Evaluating Induced Entry into a New SSDI Program with a $1 Reduction in Benefits for Each $2 in Earnings." Working draft prepared for the Social Security Administration. https://www.ssa.gov/disabilityresearch/documents/ind_entry_110501.pdf.

Vachon, Mallory. 2014. "The Impact of Local Labor Market Conditions and the Federal Disability Insurance Program: New Evidence from the Bakken Oil Boom." Paper presented at the 2014 Conference of the National Tax Association, Santa Fe, NM, November 2014. https://www.ntanet.org/wp-content/uploads/proceedings/2014/052-vachon-impact-local-market-conditions-federal.pdf.

Van Noorden, Richard, Brendan Maher, and Regina Nuzzo. 2014. "The Top 100 Papers." *Nature* 514 (7524): 550–553.

VanderWeele, Tyler J. 2011. "Principal Stratification—Uses and Limitations." *International Journal of Biostatistics* 7 (1): 1–14.

Vogl, Susanne, Jennifer A. Parsons, Linda K. Owens, and Paul J. Lavrakas. 2019. "Experiments on the Effects of Advance Letters in Surveys." In *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment*, edited by Paul Lavrakas, Michael Traugott, Courtney Kennedy, Allyson Holbrook, Edith de Leeuw, and Brady West, 89–110. Hoboken, NJ: John Wiley & Sons.

von Wachter, Till, Jae Song, and Joyce Manchester. 2011. "Trends in Employment and Earnings of Allowed and Rejected Applicants to the Social Security Disability Insurance Program." *American Economic Review* 101 (7): 3308–3329.

Vought, Russell T. 2020. *Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices*. Memo M-20-12. Washington, DC: Office of Management and Budget, Executive Office of the President.

Weathers II, R. R., and J. Hemmeter. 2011. "The Impact of Changing Financial Work Incentives on the Earnings of Social Security Disability Insurance (SSDI) Beneficiaries." *Journal of Policy Analysis and Management* 30 (4): 708–728.

Weathers II, Robert R., Chris Silanskis, Michelle Stegman, John Jones, and Susan Kalasunas. 2010. "Expanding Access to Health Care for Social Security Disability Insurance Beneficiaries: Early Findings from the Accelerated Benefits Demonstration." *Social Security Bulletin* 70 (4): 25–47. https://www.ssa.gov/policy/docs/ssb/v70n4/v70n4p25.html.

Weathers II, Robert R., and Michelle Stegman. 2012. "The Effect of Expanding Access to Health Insurance on the Health and Mortality of Social Security Disability Insurance Beneficiaries." *Journal of Health Economics* 31 (6): 863–875.

Weathers II, Robert R., and Michelle Stegman Bailey. 2014. "The Impact of Rehabilitation and Counseling Services on the Labor Market Activity of Social Security Disability Insurance (SSDI) Beneficiaries." *Journal of Policy Analysis and Management* 33 (3): 623–648.

Wehman, Paul H., Carol M. Schall, Jennifer McDonough, John Kregel, Valerie Brooke, Alissa Molinelli, Whitney Ham, Carolyn W. Graham, J. E. Riehle, and Holly T. Collins. 2014. "Competitive Employment for Youth with Autism Spectrum Disorders: Early Results from a Randomized Clinical Trial." *Journal of Autism and Developmental Disorders* 44 (3): 487–500.

Wehmeyer, Michael L. 1995. *The Arc's Self-Determination Scale: Procedural Guidelines*. Washington, DC: US Department of Education, Office of Special Education and Rehabilitative Services, Division of Innovation and Development.

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

Whalen, Denise, Gilbert Gimm, Henry Ireys, Boyd Gilman, and Sarah Croake. 2012. *Demonstration to Maintain Independence and Employment (DMIE)*. Report for Centers for Medicare & Medicaid Services. Washington, DC: Mathematica Policy Research.

Wilde, Elizabeth Ty, and Robinson Hollister. 2007. "How Close Is Close Enough? Evaluating Propensity Score Matching Using Data from a Class Size Reduction Experiment." *Journal of Policy Analysis and Management* 26 (3): 455–477.

Wilhelm, Sarah, and Sara McCormick. 2013. "The Impact of a Written Benefits Analysis by Utah Benefit Counseling/WIPA Program on Vocational Rehabilitation Outcomes." *Journal of Vocational Rehabilitation* 39 (3): 219–228.

Wing, Coady, Kosali Simon, and Ricardo A. Bello-Gomez. 2018. "Designing Difference in Difference Studies: Best Practices for Public Health Policy Research." *Annual Review of Public Health* 39: 453–469.

Wiseman, Michael. 2016. *Rethinking the Promoting Opportunity Demonstration Project*. Washington, DC: Social Security Advisory Board.

Wittenburg, David. 2011. *Testimony for Hearing on Supplemental Security Income Benefits for Children. Subcommittee on Human Resources, Committee on Ways and Means, US House of Representatives*. Washington, DC: Mathematica Policy Research.

Wittenburg, David, Kenneth Fortson, David Stapleton, Noelle Denny-Brown, Rosalind Keith, David R. Mann, Heinrich Hock, and Heather Gordon. 2018. *Promoting Opportunity Demonstration: Design Report*. Washington, DC: Mathematica Policy Research.

Wittenburg, David, Thomas Fraker, David Stapleton, Craig Thornton, Jesse Gregory, and Arif Mamun. 2007. "Initial Impacts of the Ticket to Work Program on Social Security Disability Beneficiary Service Enrollment, Earnings, and Benefits." *Journal of Vocational Rehabilitation* 27 (2): 129–140.

Wittenburg, David, and Gina Livermore. 2020. *Youth Transition*. Washington, DC: Mathematica Policy Research.

Wittenburg, David, David R. Mann, and Allison Thompkins. 2013. "The Disability System and Programs to Promote Employment for People with Disabilities." *IZA Journal of Labor Policy* 2 (4): 1–25.

Wittenburg, David, David Stapleton, Michelle Derr, Denise W. Hoffman, and David R. Mann. 2012. *BOND Stage 1 Early Assessment Report*. Report for Social Security Administration, Office of Research, Demonstration, and Employment Support. Cambridge, MA: Abt Associates.

Wittenburg, David, John Tambornino, Elizabeth Brown, Gretchen Rowe, Mason DeCamillis, and Gilbert Crouse. 2015. *The Child SSI Program and the Changing Safety Net*. Washington, DC: US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Office of Human Services Policy.

Wixon, Bernard, and Alexander Strand. 2013. "Identifying SSA's Sequential Disability Determination Steps Using Administrative Data." *Research and Statistics Notes.* No. 2013-01. Social Security Administration. https://www.ssa.gov/policy/docs/rsnotes/rsn2013-01.html.

youth.gov. n.d. "Job Corps, Program Activities/Goals." Accessed March 24, 2021. https://youth.gov/content/job-corps.

Zhang, C. Yiwei, Jeffrey Hemmeter, Judd B. Kessler, Robert D. Metcalfe, and Robert Weathers. 2020. "Nudging Timely Wage Reporting: Field Experimental Evidence from the United States Social Supplementary Income Program." NBER Working Paper No. 2785. Cambridge, MA: National Bureau of Economic Research.

Ziguras, Stephen J., and Geoffrey W. Stuart. 2000. "A Meta-Analysis of the Effectiveness of Mental Health Case Management over 20 Years." *Psychiatric Services* 51 (11): 1410–1421.