Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt

Nicole Maestas, Kathleen Mullen and Alexander Strand

M R
R C

Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt

Nicole Maestas
RAND

Kathleen Mullen
RAND

Alexander Strand
SSA

August 2012

Michigan Retirement Research Center
University of Michigan
P.O. Box 1248
Ann Arbor, MI 48104
www.mrrc.isr.umich.edu
(734) 615-0422

Does Disability Insurance Receipt Discourage Work?
Using Examiner Assignment to Estimate Causal
Effects of SSDI Receipt

Abstract

We present the first causal estimates of the effect of Social Security Disability Insurance benefit receipt on labor supply using all program applicants. We use new administrative data to match applications to disability examiners, and exploit variation in examiners' allowance rates as an instrument for benefit receipt. We find that among the estimated 23% of applicants on the margin of program entry, employment would have been 28 percentage points higher had they not received benefits. The effect is heterogeneous, ranging from no effect for those with more severe impairments to 50 percentage points for entrants with relatively less severe impairments.

1. Introduction

U.S. disability program expenditures have increased dramatically over the last several decades. As a result, the Social Security system's long-run fiscal imbalance is now an immediate problem: the Social Security Disability Insurance (SSDI) Trust Fund is expected to exhaust its assets by 2016, after which time expected tax revenues will cover only 79% of SSDI benefits (Board of Trustees, 2012).

As the SSDI caseload has grown, the employment of disabled workers has steadily declined. As Figure 1 shows, during the last two decades alone, the SSDI caseload grew three-fold while the relative employment rate of disabled workers fell by one-half. This occurred despite the facts that the Americans with Disabilities Act (ADA) created new employment protections for disabled workers, jobs became less physically demanding (Autor, Levy and Murnane, 2003), life expectancy increased (Christensen et al., 2009), and the health of the working-age population remained stable (Houtenville et al., 2009; Burkhauser and Daly, 2011).

Although demographic and economic factors have hastened the program's expansion, the consensus in the research literature is that the structure of the SSDI program itself has been a major force behind the decline in employment and concomitant program growth (Duggan and Imberman, 2009; Burkhauser and Daly, 2011). Indeed, many have attributed the long-running structural decline in employment and earnings among low-skilled men to the SSDI program, at least in part (see e.g., Parsons, 1980, 1982; Haveman and Wolfe, 1984a, 1984b; Bound, 1989; and Bound and Waidmann, 1992). More recently, Autor and Duggan (2003) show that declining labor demand for low-skilled workers during the late 1970s through 1990s, combined with a rising benefit-to-earnings replacement rate and an expansion in program eligibility in the mid-1980s, contributed to falling employment and dramatic program growth during that period.

While few would deny that the SSDI program has adversely affected employment, there is little consensus about the magnitude of SSDI's impact. Progress on this question has been hampered by the difficulty of separating the causal effect of SSDI from confounding factors that also affect the relative attractiveness of employment versus SSDI participation, such as trends in employment, earnings and health. In this paper, we solve a critical barrier to progress—the lack of exogenous variation in SSDI receipt—and present the first causal estimate of the effect of SSDI receipt on labor supply using the entire population of SSDI applicants in the present day system.

1

The precise nature of our contribution is best understood in the context of the prior literature. Bound (1989) first proposed the approach of using denied applicants as a control group for SSDI beneficiaries, as an alternative to earlier analyses that related labor supply outcomes to non-random variation in SSDI replacement rates (see e.g., Parsons, 1980). These earlier analyses, which attributed the entire decline in employment among older men to the SSDI program, overstated the impact of SSDI on labor supply because SSDI replacement rates confound benefit generosity with variation in lifetime earnings and labor force attachment (Haveman and Wolfe 1984a, 1984b; Bound 1989; Gruber 2000). Bound argued that, in observational data, the post-decision labor supply of denied applicants was an upper bound estimate of the potential labor supply of SSDI beneficiaries had they not received SSDI benefits. He contended it was an upper bound because denied applicants should have less severe impairments (and hence greater employment potential) than allowed applicants on average. He concluded that the employment of older (ages 45-64) male SSDI beneficiaries during the 1970s would have been at most around 35 percentage points higher had they never received SSDI benefits, and therefore SSDI accounted for only about one-quarter of the historic decline in employment among older men.

Parsons (1991) argued that Bound's estimate was not an upper bound on the *total* program effect if the application process also reduces the labor supply of applicants. All applicants must demonstrate that they are not working before they may apply, and initially denied applicants must remain out of the labor force to maintain eligibility for benefits while they pursue appeals or reapply. This initial withdrawal from and time spent out of the labor force while seeking benefits may diminish human capital, making it harder for all applicants to re-enter the labor force.[1] Moreover, if the application process affects the post-decision labor supply of ultimately allowed and denied applicants differentially, then Bound's approach would not even give an upper bound on the effect of benefit receipt. More generally, Bound's approach would understate the employment potential of SSDI beneficiaries to the degree that denied applicants have lower employment potential than allowed applicants.[2]

---

[1] New evidence from Autor, Maestas, Mullen and Strand (2011) suggests that the SSDI program may indeed affect employment outcomes through channels other than benefit receipt, such as processing time.

[2] This is certainly possible. Allowance decisions for applicants with less severe impairments explicitly take account of vocational factors that are inversely correlated with lifetime labor force attachment such as age, education, and work experience, and later we show that denied applicants in fact have lower prior labor force attachment than allowed applicants.

Von Wachter, Manchester and Song (2011) applied Bound's comparison approach to observational data for the 1980s and 1990s. Using administrative records that permit the tracking of appeals outcomes, they conclude that the overall employment potential of the SSDI caseload has risen significantly since the 1970s owing to an increase in beneficiaries who are younger and who have nonterminal impairments such as mental health and musculoskeletal conditions. Still, Von Wachter, Manchester and Song lacked an exogenous source of variation in SSDI receipt, and thus were unable to resolve the fundamental issue of causal magnitude.

Two other studies improved on the observational approach by using policy variation in initial allowance rates to mitigate the confounding with lifetime earnings and impairment severity. Gruber and Kubik (1997) used differential decreases in state-level initial allowance rates during the late 1970s that resulted from federal policy changes aimed at reducing program growth. Their reduced form estimates imply that labor force participation among older men would have been 8.9 percentage points greater had they not been initially allowed benefits.[3] Gruber and Kubik did not estimate the "first stage" relationship between the state-level *initial* allowance rate and an individual's *ultimate* allowance probability (which, again, would account for appeals/reapplication outcomes) so it is not possible to scale up the reduced form effect to find the implied causal effect of SSDI receipt on labor supply. Chen and van der Klaauw (2008) used variation in initial allowance rates arising from program rules that relax the vocational eligibility standard discontinuously at ages 45, 50 and 55 using survey data covering the 1990s. Their estimates are difficult to compare to the prior literature because of sample differences; they necessarily analyzed only vocational applicants (those who did not qualify on the basis of medical criteria alone), but also included SSI applicants and women. Their preferred estimate suggests that the employment of older vocational beneficiaries would have been 20 percentage points higher in the absence of initial allowance, but like Gruber and Kubik (1997) their estimates are reduced form estimates because they could not track appeals outcomes in their data. If appropriately scaled up by the first stage relationship between initial and ultimate allowance, their estimate implies substantial work capacity among vocational beneficiaries, perhaps even *more* capacity than Bound's estimate implies.[4]

---

[3] Gruber and Kubik (1997) report that a 10% increase in the denial rate led to a 0.48 percentage point decrease in labor force participation. The mean denial rate in 1977 was 53.8%.

[4] Chen and van der Klaauw (2008) also apply Bound's comparison approach to their data and find that their quasi-experimental estimates are consistently *larger* than the upper bound observational estimate. This implies that a

Our study advances this literature in several ways. Our first contribution is the development of a research design that had not previously been used in this literature, but which enables us to estimate the causal effect of SSDI receipt on labor supply along a margin of major policy interest—program entry. To do this, we compare the subsequent labor supply of otherwise similar applicants who were allowed or denied benefits only because their applications were randomly assigned to disability examiners with different allowance propensities at the initial point in the disability determination process. This assignment mechanism generates exogenous variation in decision outcomes that is unrelated to unobserved impairment severity or labor force attachment. We implement this research design in a unique administrative data set that has been underused by researchers and contains disability examiner identification codes for the universe of SSDI applications in a given year. Because the disability determination process allows for multiple levels of appeal, we employ an intention-to-treat framework (Angrist, Imbens and Rubin, 1996) in which we use an examiner's allowance propensity in the initial determination stage as an instrumental variable for the *ultimate* allowance decision for a given application. We identify ultimate allowance decisions (taking account of appeals outcomes) by linking the applications data to administrative data recording SSDI benefit receipt up to 6 years after the initial decision. We then link our data to administrative earnings records and examine labor supply outcomes 2 to 4 years from the date of the initial determination. By waiting 2 to 4 years to measure labor supply outcomes, we address the likelihood that ultimately denied applicants kept their earnings low during the appeals process.

We find that the employment rate of beneficiaries on the margin of SSDI entry in 2005 and 2006 (23% of all applicants) would have been on average 28 percentage points higher two years later if they had never received SSDI benefits. Employment above SSA's threshold defining substantial gainful activity would have been 19 percentage points higher and annual earnings would have been $3,781 higher (including the zeros). Although the estimate of employment capacity is large, our estimate of earnings capacity indicates that the average work

comparison of allowed and denied applicants in observational data does not necessarily yield an upper bound on the potential labor supply of beneficiaries. Rather, this pattern suggests that for older vocational applicants (39 percent of all applicants in their sample) unobserved differences in labor force attachment between allowed and denied applicants may dominate unobserved differences in impairment severity. This is a reasonable interpretation since vocational applicants are more likely to have been economically induced to apply for SSDI benefits and allowance decisions for vocational applicants depend on factors that are negatively correlated with labor force attachment. Chen and van der Klaauw's use of exogenous variation in allowance rates naturally resolves the downward bias in the observational estimate.

capacity of this group is far below their pre-disability earnings levels. Still, our results show that a subset of SSDI beneficiaries either retains or recovers some degree of work capacity in the years immediately following their initial decision.

Concurrent with our study, French and Song (2011) use variation in the propensity of administrative law judges (ALJs) in the second stage of the appeals process to estimate the labor supply effect of SSDI receipt. They find that the employment rate of applicants allowed at the ALJ hearing level would be 26 percentage points higher if they had instead been denied. Our framework enables us to compare their estimates to ours and explain why, despite important differences in the composition of the two samples (i.e., plaintiffs in ALJ-level appeals cases vs. all initial applicants), it is reasonable to expect the two approaches to produce similar estimates.[5]

Our research design facilitates two additional contributions to the literature. First, we estimate that 23% of SSDI applicants are on the margin of program entry in the sense that whether they ultimately receive benefits depends on their initial examiner assignment. The size and characteristics of this group are of significant policy interest, since these applicants would be most affected by policy changes addressing disability determinations. We find that this group on the margin of program entry disproportionately includes younger applicants, applicants with low earnings histories, and applicants with mental impairments. In addition, we calculate that the program entry margin (accounting for appeals outcomes) spans the 20th through 43rd percentiles of the distribution of unobserved severity among all applicants.

Second, our continuous instrumental variable offers a first opportunity to test for heterogeneous treatment effects. We find that the causal effect of SSDI receipt on employment is not the same across all beneficiaries. Rather, employment capacity rises as high as 50 percentage points for marginal applicants with the least severe impairments and falls to zero for those with the most severe impairments. We also estimate the distribution of treatment effects among applicants on the margin of program entry.

2. Background on SSDI

The SSDI program defines disability as the "inability to engage in substantial gainful activity (SGA) by reason of any medically determinable physical or mental impairment(s) which

---

[5] One important difference is that French and Song use data on ALJ decisions from the 1990s, whereas we use data on initial decisions from the mid-2000s. To the extent that the composition of applicants has changed over time, this could affect the underlying work capacity available in the SSDI population.

can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months." Activity is considered "substantial" if it involves significant physical and/or mental exertion, and it is considered "gainful" if it is performed for pay or profit (whether or not profit is actually realized). SSA operationalizes this definition by setting an earnings threshold – currently $1,010 per month – over which individuals are said to be engaging in SGA and are therefore disqualified from participating in the program. The SGA threshold impacts both program entry (through eligibility) and the labor supply of current beneficiaries (whose benefits are suspended if they earn more than the SGA threshold once they have completed a trial work period).

Individuals apply for SSDI benefits at their local field office, which screens out those who are not currently insured or who are engaging in SGA.[6] These are labeled "technical denials," and do not receive further review.[7] The remaining applications are forwarded to a state Disability Determination Services (DDS) office, where cases are assigned to disability examiners for review.[8] During our sample period (2005-2006), the vast majority of initial claims for adults were allocated to examiners using a computer program developed by Iron Data LLC, although some states used their own program to process cases.[9] We conducted telephone interviews with managers of the ten largest DDS offices (all of which use the Iron Data software) as well as two offices that used their own software. These twelve offices processed 40% of the claims in our sample in 2006. The interviews revealed that nearly all cases were automatically assigned to examiners by computer program, which assigned new applications to examiners based on availability.[10] The managers variously described this process as a "round robin," "next in line," "rotate to the back," or "equal distribution" system.

---

[6] Roughly half of SSDI applicants also concurrently apply for benefits from the Supplemental Security Income (SSI) program, which provides additional income to individuals with limited assets. The medical and vocational criteria are the same under both programs.

[7] The field office also screens out applications from individuals who do not qualify for SSDI based on their work history. Generally, to be eligible for SSDI one must have earned at least 40 work credits (or "covered quarters"), 20 of which must have been earned in the last 10 years. However, these restrictions are relaxed for younger individuals.

[8] There are 114 DDS offices. Two-thirds of states (including Washington, D.C.) have only 1 DDS office and 47 states have 4 or fewer regional DDS offices. States with more than 4 DDS offices are: New York (5), Missouri (6), Florida (10) and California (10). There are also 7 offices handling claims in U.S. territories.

[9] In January 2011, SSA announced that all DDS offices would switch to a new (common) case management system (to be provided by Lockheed Martin with Iron Data LLC acting as subcontractor).

[10] Even prior to the era of computer assignment, there is evidence that cases were randomly assigned to examiners. See, for example, Lewin Group, Inc. (2001), which conducted site visits in 1999 as part of a study on the quality of disability determinations.

Two types of exceptions were noteworthy in our interviews with DDS managers.[11] First, in all DDS offices a small fraction of cases are sent by the field offices with high-priority flags for expedited handling, and these are sometimes manually assigned to examiners who specialize in high-priority cases. The only high-priority cases during 2005-2006 were for allegations of terminal illness (TERI).[12] TERI claims are characterized by impairments that cannot be reversed and are expected to end in death, and have very high allowance rates. An allegation of certain conditions by the claimant will result in the claim being marked TERI by the field office.[13] Second, in some (although not all) DDS offices, the applicant's affected body system is taken into account in assignment of cases to newly hired examiners as part of their training.[14]

All screening personnel are required to follow a five-step review process in order to determine whether applicants are unable to perform SGA given the severity of their impairment and/or vocational background. (See Figure A-1 in the appendix for a visual overview of the process.) Denials are issued to applicants engaging in SGA (step 1—performed at the field office) and to applicants with impairments that are obviously temporary or non-severe (step 2). Next, "medical" allowances are made to individuals with specific impairments deemed severe enough to warrant allowance into the program regardless of vocational background (step 3). The remaining applicants are evaluated to determine whether they have the residual functional capacity to perform any of their past jobs (step 4) or any job for which they are qualified in the national economy given their vocational background (step 5). Denials may be issued at either of these two steps, and "vocational" allowances are issued to qualified applicants at step 5. Vocational allowances are made in consultation with a medical-vocational grid that provides guidance for allowance decisions based on residual functional capacity, age group, education and

---

[11] Other exceptions emerged but these seemed unlikely to be concentrated among examiners. For example, high profile cases or cases involving conflicts of interest with DDS personnel could be reassigned manually.

[12] Other types of high-priority cases were introduced after our sample period. Since 2007, SSA has introduced the Quick Disability Determination (QDD) program, which identifies electronically via a predictive model those claims that have a high probability of allowance and for which evidence is expected to be easily and quickly obtained, and the Compassionate Allowance (CAL) program, which identifies cases likely to qualify under the medical listings. Similarly, in 2008 SSA began enhanced processing of military service casualty claims ("Wounded Warriors") using the expedited procedures.

[13] These conditions include: chronic dependence on a cardiopulmonary life-sustaining device; awaiting certain organ transplants; chronic pulmonary or heart failure requiring continuous home oxygen and unable to care for personal needs; any malignant neoplasm (cancer) which is metastatic, stage IV, persistent or inoperable; an allegation or diagnosis of ALS, AIDS, or one of eight types of cancer; or, comatose for 30 days or more.

[14] In many offices, less experienced examiners handle only initial determinations, whereas more experienced examiners also handle more complex types of cases such as reconsiderations or continuing disability reviews (CDRs).

type of work experience (i.e., skilled, unskilled).[15] Disability examiners are not medically trained but may consult with a medical consultant (a physician or psychologist) in order to assess an applicant's residual functional capacity, or ability to work given their physical and/or mental impairments.

Denied applicants can appeal their initial determination within 60 days by applying for reconsideration by the original DDS office.[16] Applicants denied at reconsideration have an additional 60 days to file an appeal to have an administrative law judge (ALJ) review their case. The ALJ must consider the application using the same steps in the same order as the initial determination, but the applicant may present new information. Applicants denied by an ALJ have additional opportunities to appeal to the SSA Appeals Council (AC) and finally to Federal Court. In our data, roughly one-third of applicants are allowed in the initial determination (with about 60% of these receiving vocational allowances), and just under two-thirds of applicants are ultimately awarded SSDI benefits within six years of their initial determination. An audit study performed by the Office of the Inspector General (OIG) (2008) estimated that in 2006 the average processing time for cases in the initial determination phase was 131 days, or just over one-third of a year. Average (cumulative) processing times for cases reaching the appeals phases were 279 days (0.76 years) for reconsideration, 811 days (2.22 years) for ALJ, 1,053 days (2.88 years) for AC and 1,720 days (4.71 years) for Federal Court (OIG 2008). Just under one-third of cases made it to the ALJ level, where approximately 61% of initial denials were overturned (GAO 2004). Fewer than 5% of cases progressed to the AC level and fewer than 1% of cases progressed to Federal Court.

Individuals who apply for SSDI benefits must stop working or reduce their earnings below the SGA threshold for a period of five months before they are entitled to receive benefits, or until the allowance decision is made (whichever comes later).[17] The SSDI benefit formula is the same progressive formula used to calculate Social Security benefits based on past average indexed monthly earnings, and does not depend on the type or severity of the individual's

---

[15] Practically, the grid generates increases in the allowance rate at ages 45, 50, and 55, which was exploited in the regression discontinuity design of Chen and van der Klaauw (2008).

[16] Since 1999, ten "prototype" states have eliminated the reconsideration step of the appeals process; they are: Alabama, Alaska, California (Los Angeles North and Los Angeles), Colorado (West), Louisiana, Michigan, Missouri, New Hampshire, New York and Pennsylvania.

[17] In addition to the five-month waiting period for SSDI benefits, there is a nineteen-month waiting period for healthcare coverage under Medicare. Neither waiting period exists for SSI, which immediately entitles its beneficiaries to cash benefits and Medicaid.

impairment. The average monthly benefit was $1,129 in 2010 (Social Security Administration, 2010, Table 36). Once benefits commence, beneficiaries begin a nine-month Trial Work Period (TWP) which allows individuals to test their ability to return to work by relaxing the restriction that earnings may not exceed the SGA threshold. The TWP is followed by a three-month Grace Period before individuals earning above the SGA threshold have their benefits suspended. Thus, SSDI beneficiaries may engage in SGA for up to twelve months (not necessarily consecutively) while receiving their full benefits without any penalty. Over the next three years, during an extended period of eligibility (EPE) benefits are paid for months in which earnings are below SGA, and not paid when earnings are above SGA.

Upon reaching the Social Security Full Retirement Age, SSDI benefits are automatically converted to Social Security retired worker benefits and the SGA earnings restriction is lifted. Very few beneficiaries exit the program for a reason other than death (35% in 2009) or automatic conversion to retired worker benefits at the Full Retirement Age (54%). In 2009, only 8% of worker beneficiaries' benefits were terminated because they no longer qualified for benefits; of those, 39% failed a medical review (a Continuing Disability Review, or CDR) whereas 61% were found to be earning more than the SGA threshold (Social Security Administration, 2010, Table 50).

3. Data and Caseload Characteristics

We make use of a unique workload management database called the Disability Operational Data Store (DIODS). The DIODS contains the universe of SSDI applications, and according to Social Security Online, is considered "the SSA definitive data store for disability claims for state agency workload management purposes."[18] The DIODS contains alphanumeric codes that identify disability examiners uniquely within DDS and allow us to construct a measure of each examiner's propensity to allow SSDI applicants.[19] We use data on all initial medical determinations (that is, excluding technical denials) made between January 1, 2005, and December 31, 2006. We construct our sample by restricting to disabled workers only between

---

[18] http://www.socialsecurity.gov/disability/data/ssa-sa-mowl.htm

[19] Unfortunately, we do not observe any examiner characteristics. For example, we cannot distinguish between full- and part-time examiners, nor can we observe examiner tenure.

the ages of 18 and 64 who were assigned to examiners handling at least 30 cases in the observed time frame.[20] We further exclude applicants who were previously SSDI or SSI beneficiaries.[21]

We link the DIODS data to SSA's Master Beneficiary Record (MBR) in order to identify which applicants ultimately received SSDI benefits (by the end of 2010). This allows us to identify SSDI receipt at least four years and up to six years after the initial determination. In order to identify applicants who appealed an initial denial to the administrative law judge (ALJ) level, we also linked our sample with an extract from the Case Processing and Management System (CPMS), which includes information on hearings conducted through June 2011.[22] Because we observe benefit receipt four to six years after the initial decision, we conservatively estimate that we correctly classify benefit receipt for at least 99.4% of applicants in our sample.[23]

To examine past and subsequent labor supply, we link to administrative annual earnings records between 1995 and 2009 from SSA's Detailed Earnings Record (DER). The DER contains uncapped earnings from box 5 (Medicare wages and tips) of individuals' W-2 forms. We aggregate across all earnings reports to measure earnings at the individual level. This gives us annual earnings up to nine years before and up to four years after the initial decision.[24] We use three measures of labor supply: 1) employment– defined as earning more than $1,000 (in 2008 dollars) in a given year;[25] 2) employment above SGA (where the SGA threshold is that defined by SSA in a given year); and 3) annual earnings. All earnings amounts are presented in 2008 dollars, and include those with zero earnings. We examine labor supply at two to four years after the initial decision to permit most of the denied applicants to complete the appeals process.

---

[20] One DDS manager indicated that trainees all have the same examiner code in the DIODS for her DDS. Although we do not believe this practice to be widespread, we eliminated cases assigned to examiner codes associated with more than 900 cases in the two year decision period in order to avoid this potential problem. This additional restriction results in the exclusion of 24 examiners and 37,631 cases.

[21] This eliminates claims assigned to specific examiners associated with the "Special Disability Workload," which resulted from an outreach initiative by SSA to determine whether recipients of SSI payments had worked enough to become insured for SSDI (OIG 2006).

[22] Unfortunately, we do not observe reconsideration-level appeals, or appeals beyond the ALJ level.

[23] This is based on the fact that 2% of applications result in benefit receipt despite being denied at the ALJ level; the median and 90th percentile of cumulative processing times among these applicants are four and five years, respectively (see Table 2 in Autor, Maestas, Mullen and Strand, 2011). Thus, we misclassify at most 0.6% of outcomes (=0.01*0.1 of 2005 applications plus 0.01*0.5 of 2006 applications) as denials rather than allowances.

[24] Earnings are measured between January 1 and December 31 of a given calendar year. Thus, an applicant whose decision was made in December of 2005 will have had 11 fewer months elapse between his decision and outcome year (say, 2007) than an applicant whose decision was made in January of 2005, even though we refer to the outcome as being measured two years later for both applicants.

[25] Because any positive earnings result in a W-2 report, we observe many individuals with very small annual earnings. We impose the $1,000 earnings threshold to restrict our definition of employment to only "meaningful" participation in the labor market.

Finally, we link the sample to SSA's Numerical Identification System (Numident) file, which contains dates of death for individuals who died during our analysis period. We restrict the analysis samples to individuals who were alive through the end of the calendar year in which earnings are observed.[26]

Table 1 presents summary statistics for our two-year analysis sample, overall and by case disposition.[27] The sample consists of just over 2.1 million applications. These cases were adjudicated by approximately 8,700 DDS examiners, for an average caseload of roughly 250 initial determinations in the two-year period, or 125 per year.[28] Table 1 divides the sample into four groups: 1) initially allowed, 2) initially denied, but allowed on appeal (or through reapplication), 3) initially denied, and denied on appeal, and 4) initially denied, and not observed to appeal.[29] Approximately 35% of applicants denied at the initial determination level do not appeal their decision. However, among those who appeal, the success rate is fairly high – fully three-quarters of these denials are eventually overturned. This underscores the importance of using an intent-to-treat framework, and also is the reason why previous estimates based only on variation in initial allowance rates are reduced form estimates, and not estimates of the effect of benefit *receipt*.

Fifty-nine percent of applications are for one of two types of impairments: musculoskeletal (38%) or mental disorders (21%). These percentages are virtually the same among all allowances, although musculoskeletal (mental) cases are less (more) likely to be allowed at the initial determination level than on appeal. Fewer than 1% of all applications are cases of alleged terminal illness, and these cases have very high initial allowance rates, representing just over 2% of initial allowances.[30] Among the initially denied applicants, the three

---

[26] SSDI applicants have relatively high rates of mortality, especially among beneficiaries. Livermore, Stapleton and Claypool (2009) estimate that 13% of SSDI beneficiaries die within two years of entering the program.

[27] These will differ from published statistics since we impose some additional sample restrictions, such as excluding high mortality cases (those who died in the two years following the initial determination) as well as applicants assigned to examiners with less than 30 initial determinations.

[28] This number actually underestimates the number of initial determination decisions per year since it imposes our sample restrictions. In addition, note that initial determinations represent only about a third of all cases decided by examiners, which also include reconsiderations, continuing disability reviews, and SSI determinations, as well as SSDI claims filed by dependents.

[29] We identify individuals who appeal by matching initial applications to cases presented to administrative law judges. Thus, group 4 includes some individuals who appealed for reconsideration, but upon denial at that stage, declined to appeal further. It also may contain denied applicants who submitted a new application instead of appealing, but whose subsequent application was also denied.

[30] While we do not have a direct measure of terminal illness cases in our data set, we construct an indicator for alleged terminal illness using detailed diagnosis codes corresponding to the impairments listed in Section 2. Our

groups are ordered in terms of impairment severity; applicants who did not appeal were more likely to be denied in earlier steps of the initial review process than those who appealed but were subsequently denied, who were more likely to be denied earlier than those who successfully appealed. Ultimately allowed applicants also tend to be older and have higher pre-onset earnings than ultimately denied applicants.

Figure 2 provides a first descriptive look at the data by plotting the employment rate of 2005 applicants up to nine years before and four years after their initial determination.[31] We restrict the sample to 2005 applicants because outcomes for 2006 applicants are only observed up to three years after their initial determination.[32] From Figure 2 we see that before the initial determination, the employment rate of applicants allowed on appeal is very similar to that of applicants who were initially allowed – above 90% five years before the initial decision. From this peak, employment declines steadily in the years leading up to the initial decision, then drops sharply in the decision year. By three years after the decision the two groups are indistinguishable with employment rates around 13%.

In contrast, ultimately denied applicants have lower employment rates to begin with – 80% five years before the initial determination – and significantly higher (albeit reduced) employment rates (around 50%) after the initial determination.[33] Although denied applicants who appeal (unsuccessfully) and those who do not appeal have similar pre-decision employment rates, their post-decision employment differs substantially. Specifically, only about 40% of individuals who appealed unsuccessfully are working three years after their initial determination. Given average cumulative processing times reported by the OIG (2008), very few of these applicants are likely to be still awaiting a decision (and hence suppressing labor supply) 3-4 years after the initial decision. Rather, the gap between the two types of denied applicants likely reflects selection into the appeals process; that is, individuals with less severe impairments are

estimated fraction of terminal illness cases is consistent with administrative counts and internal SSA studies of mortality rates among high-priority claims.

[31] Figures A-2 and A-3 in the appendix present the same graphs for the other two outcome measures: earning above the SGA threshold and mean earnings.

[32] Consistent with our later empirical results, we do not condition on a balanced panel for Figure 2 (that is, individuals living up to four years after the initial determination). However, the figures and estimates are similar with and without the restriction of a balanced panel.

[33] From Table 1 we calculate that approximately 68% (=0.24/(0.11+0.24)) of those ultimately denied benefits did not appeal the initial determination.

more likely to opt out of the appeals process, whether by choice or because they fail to secure a lawyer willing to represent their case.

Similarly, differences in prior employment between ultimately allowed and denied applicants are consistent with selection on non-health factors into the applicant population. If individuals with low labor supply prior to disability onset have lower opportunity costs of applying, they will be more likely to apply conditional on a given level of health, and as a result prior labor supply and health will be negatively correlated among applicants. The relationship between prior labor market history and ultimate disposition of the case becomes even more pronounced when examining the percent of applicants engaging in SGA (i.e., earning more than $11,760 per year in 2009) (see Figure A-2) and average earnings (Figure A-3).[34]

4. Empirical Strategy

The goal of this paper is to estimate causal models of labor supply of the following form:

$$y_i = X_i\beta + \gamma DI_i + \upsilon_i, \tag{1}$$

where y_i is earnings (or employment) of applicant i, X_i denotes observed characteristics (e.g., age, type of impairment) that may influence labor supply, $DI_i = 1$ if the applicant was awarded SSDI benefits (i.e., is observed to be a SSDI beneficiary up to six years after the initial determination), and υ_i is an error term. In observational data, inference is hampered if some unobserved characteristic, such as severity of the impairment, impacts both labor supply and SSDI allowance. For instance,

$$y_i = X_i\beta + \gamma DI_i - s_i + \varepsilon_i,$$

where s_i denotes unobserved severity, which can be thought of as (the unobserved part of) the earnings loss associated with the individual's impairment, and which is uncorrelated with any remaining idiosyncratic element ε_i. Then in the regression above $\upsilon_i = -s_i + \varepsilon_i$, and if $E[s_i \mid DI_i] \neq 0$, ordinary least squares (OLS) regression gives a biased estimate of the average treatment effect, γ. In particular, OLS estimates $\gamma - [E[s_i \mid DI_i = 1] - E[s_i \mid DI_i = 0]]$. As observed by Bound (1989), if $\gamma < 0$ and if the unobserved characteristic is positively correlated

---

[34] Note that, because of the earnings restriction, these are unlikely to be the same beneficiaries engaging in SGA in more than one year.

with SSDI receipt, OLS overestimates the magnitude of the coefficient on *DI* and provides an upper bound on the labor supply effect of benefit receipt.

From SSA's point of view, in an ideal world SSDI benefits are awarded to individuals whose potential earnings – in the absence of SSDI benefits – are less than the SGA threshold:

$$X_i \beta - s_i + \varepsilon_i < SGA.$$

In practice, however, cases are assigned to disability examiners who have imperfect information, and so the assignment rule becomes based on the contrast:

$$X_i \beta - \hat{s}_{ij} < SGA,$$

where $\hat{s}_{ij}$ denotes the estimate by examiner *j* of the severity of individual *i'*s impairment. This estimate is a function of both the individual's impairment severity – which the examiner observes in greater detail than the econometrician, through medical records and test results[35] – and characteristics of the examiner assigned to the case, such as previous experience or personal perceptions and tastes. Let

$$\hat{s}_{ij} = s_i + \sigma_j,$$

where σ_j denotes a systematic component of examiner judgment that leads some examiners to over- or underestimate severity of applicants on average. Then the assignment rule becomes

$$DI_i = 1(s_i > X_i \beta - SGA - \sigma_j), \tag{2}$$

which implies that examiner *j*'s allowance propensity conditional on applicant characteristics is

$$P_j = P(\sigma_j \mid X_i) = 1 - F(X_i \beta - SGA - \sigma_j), \tag{3}$$

where F denotes the cumulative distribution function for unobserved severity. Equations 2 and 3 show that high-σ_j examiners systematically overestimate severity, resulting in lower allowance thresholds and higher propensities to allow applicants conditional on applicant characteristics. This implies a natural identification strategy for estimating the labor supply effect of SSDI benefit receipt. In particular, we implement an instrumental variables (IV) estimation

---

[35] Note that we do not observe any medical records in our data set, only the type of impairment. In addition, in some cases the examiner may communicate with the treating doctor to clarify an aspect of an applicant's file, and these communications may not be part of the applicant's more detailed record.

strategy where we instrument for SSDI receipt in Equation 1 using the assigned examiner's propensity P_j.[36]

In order to avoid biasing measurement of examiner propensities using the applicant's own decision, we construct the following individual-specific instrument for SSDI receipt:

$$EXALLOW_{ij} = \frac{n\_allowed_j - 1(DI_i = 1)}{n\_cases_j - 1}.[37]$$

Intuitively, EXALLOW measures the allowance rate of examiner *j*, to whom applicant *i* was assigned, for all cases *except* for the case of applicant *i* himself. Thus, conditional on observed characteristics and assuming conditional random assignment, which we discuss below, EXALLOW should be positively correlated with the applicant's own award decision only if there exists an underlying examiner-specific threshold for allowance decisions.

Table 1 reports the mean (0.36) and standard deviation (0.10) of EXALLOW. The mean absolute deviation is 0.08. The range of EXALLOW runs virtually the entire length of the unit interval: 0.02 to 1. However, only a few examiners have such extreme allowance rates: the 1st and 99th percentiles of EXALLOW are 0.17 and 0.64, respectively.[38] Figure 3 presents smoothed histograms at the examiner level of examiners' deviations from the mean initial allowance rate in their DDS office, unadjusted and regression-adjusted for differences in case mix. Case controls include the fraction of cases in each of nine age bands, fourteen body system codes, alleged terminal illness, 3-digit zip code and decision month, as well as a variable measuring average prior earnings of the set of applicants assigned to a given examiner. Adjusting for case mix reduces variation in initial allowance rates, but there is still significant variation remaining (the standard deviation is 0.06, compared with 0.10 unadjusted).

---

[36] The estimation strategy we employ is similar to that used by Kling (2006) to examine the effect of incarceration on labor supply and earnings, Doyle (2007, 2008) to examine the effects of foster care placement on juvenile delinquency and adult crime, and Perry (2008) to examine the effect of treatment of maternal depression on management of children's asthma.

[37] An alternative approach would be to regress allowance decisions on a full set of examiner fixed effects in the first stage. A disadvantage of this approach is that IV is then susceptible to a "many weak instruments" problem (Stock, Wright and Yogo 2002). In addition, our formulation of the examiner allowance propensity as a continuous instrumental variable has the advantage of allowing us to compute marginal treatment effects, as can be seen below.

[38] Despite the fact that we condition on examiners with caseloads of 30 or more, one might be concerned that examiners with relatively few observations will tend to have very high or very low allowance rates because they are noisier. We explored this possibility by applying a Bayesian "shrinkage" estimator to EXALLOW (see, e.g., Kane and Staiger 2008) and estimating our results using this "corrected" instrument. The new instrument had a range of 0.14 to 0.75. Both the first and second stage (labor supply) estimates were slightly higher using this alternative instrument, but not significantly so, and the patterns in the coefficients remained the same.

Two key assumptions underlie our empirical strategy. First, in order for EXALLOW to be a valid instrument for SSDI receipt, applicants' assignment to DDS examiners must be uncorrelated with unobserved characteristics such as impairment severity conditional on observed characteristics. This amounts to an assumption of *conditional* random assignment to DDS examiner *within a DDS*. That is, at most, examiners may specialize in a particular type of impairment (e.g., mental disorders) or age group, but within this type, examiners do not further specialize in cases of either low or high severity. As discussed previously, applicants are assigned to examiners using a rotational process that is unrelated to applicant characteristics except in certain cases (alleged terminal illness or body system), which we observe in our data set. We provide empirical evidence consistent with conditional random assignment in Section 5. Because applicants are assigned to examiners regionally, controlling for DDS office is crucial to our identification strategy; otherwise, variation in EXALLOW could reflect cultural differences in disability determination standards across DDS offices, as well as differences in local applicant populations arising from differences in wage and employment opportunities in local labor markets. We therefore include DDS office fixed effects in all empirical specifications.

A second assumption that is critical to the causal interpretation of our IV estimates is the assumption that examiners' award propensities affect applicants' chances of SSDI receipt in the same way (monotonicity). In other words, we assume that cases allowed by "strict" examiners (those with low σ_j's) would also have been allowed by "lenient" examiners (those with high σ_j's), and that cases denied by lenient examiners would also have been denied by strict examiners. While not strictly testable without a well-designed audit study, this assumption implies that denials by strict examiners should be overturned on appeal at a higher rate than denials by lenient examiners, and thus the instrument should have a weaker effect on individuals assigned to strict examiners.[39]

Figure 4 provides a graphical overview of our estimation strategy, where the left panel presents the visual first stage, and the right panel presents the visual reduced form. We separately plot smoothed SSDI receipt and employment rates by examiner's residualized (regression-adjusted) initial allowance rate (with dashed lines indicating 95% confidence interval bands)

---

[39] A second implication of the monotonicity assumption is that examiners who are "strict" on one type of case, say musculoskeletal impairments, are also strict on any other type of case, say mental disorders. Below we provide empirical evidence that allowance rates for different body system codes are positively correlated within examiner.

estimated via local quadratic regression. Consistent with our earlier estimates, SSDI receipt is increasing in the residualized initial allowance rate. Notably its *slope* is also increasing in the residualized initial allowance rate. This is consistent with the implication of monotonicity that individuals initially denied by strict examiners are on average less healthy than individuals initially denied by lenient examiners, and thus more likely to obtain benefits on appeal or through reapplication. Whereas SSDI receipt rises with the examiner's allowance propensity, the right panel shows that the employment rate of applicants two years after the initial decision falls, from just under 30% among individuals assigned to very strict examiners to just under 24% among individuals assigned to very lenient examiners.

5. First Stage Estimates: Initial Disability Determinations and SSDI Receipt

In this section we present our first stage estimates and use them to draw inferences about the size and characteristics of the population of applicants on the margin of program entry. Beginning with the first stage estimates themselves, Table 2 presents the coefficient on EXALLOW in the applicant-level regression of SSDI receipt on EXALLOW, with and without covariates. We present results for 2005 and 2006, separately as well as pooled. All models include DDS indicators to account for stratification of examiners across DDS offices. We display t-statistics in parentheses, where robust standard errors are computed and clustered by DDS examiner. Column 1 shows the first stage coefficient on EXALLOW from a regression with no additional covariates. In both years, a 10 percentage point increase in initial examiner allowance rate leads to an approximately three percentage point increase in the probability of ultimately receiving SSDI.

Adding covariates sequentially to the regression allows us to indirectly test for random assignment on the basis of observable characteristics because only covariates that are correlated with EXALLOW will affect the estimated coefficient on EXALLOW when included. Based on our interviews with DDS managers (see Section 2), we expect the additions of the body system and terminal illness indicators to potentially affect the coefficient on EXALLOW, since they are case assignment variables, but no other variables should affect the coefficient. The coefficient on EXALLOW falls from 0.29 to 0.24 with the addition of body system codes, and is not significantly affected by the addition of any other variables, including the TERI flag. Thus, our

results are consistent with random assignment of applicants to examiners within DDS office, conditional on body system code and alleged terminal illness.[40]

Column 7 shows that conditional on observables, examiner allowance rates are strongly correlated with individual benefit receipt: our first stage coefficient of 0.23 implies that a 10 percentage point increase in the examiner allowance rate is associated with a statistically significant (p<0.0001) 2.3 percentage point increase in the probability that an applicant ultimately receives SSDI benefits.[41]

5.1. Size of the Population of Applicants on the Margin of Program Entry

Our first stage coefficient also gives an estimate of the fraction of applicants on the margin of program entry; that is, those applicants who might have received a different allowance decision had their case been assigned to a different initial examiner in the same DDS office. More precisely, in the case of a binary treatment (i.e., SSDI receipt), the size of the marginal population is the first stage coefficient times the range of initial allowance rates (in this case 1, see Section 4). Thus, we estimate that 23% of applicants in 2005 and 2006 represent cases on the margin of program entry.[42, 43] Given an initial allowance rate of 33%, this implies that 7.6% of applicants (or 12% of incoming beneficiaries) are allowed only because of the examiner to whom

---

[40] We also experimented with a different measure of initial allowance rate to test the implication of the monotonicity assumption that generic allowance rates can be used to instrument for any type of case. For this measure, we constructed the initial allowance rate leaving out all cases with the same body system code as the applicant (instead of just the applicant's own case). Table A-1 presents these results. For all impairments but one ("special/other" cases, around 4% of the sample), this alternative measure of EXALLOW is positively and significantly associated with increased SSDI receipt. (We replicated our analysis of labor supply effects dropping this group and obtained virtually identical results.) If we alternatively constructed EXALLOW using only cases of the same impairment type, then monotonicity would only have to hold within impairment type. However, this approach results in loss of precision and a selected sample due to small samples within impairment type at the examiner level. Table A-2 presents estimates of the first and second stages using this alternative instrument.

[41] Recall that in a model with one endogenous regressor and one instrument the t-statistic squared is equal to the incremental F statistic (in this case F=1592).

[42] It is important to note that this estimate accounts for the appeals process, which dampens the effect of initial examiner assignment considerably. In fact, we estimate the fraction of cases whose *initial* decision depends on their examiner to be 60%. Previous studies sponsored by SSA also document variation in disability evaluations (e.g., Nagi, 1969, Gallichio and Bye, 1981); differences in methodology and caseload composition make it difficult to compare those studies to ours.

[43] Although 23% of applicants *could* have received a different allowance decision had they been assigned to a different examiner, fewer *would* have received a different disposition if they had been assigned to a different examiner (in the same DDS office). This is because most examiners have initial allowance rates close to the average for their DDS office. For example, if all applicants were reassigned to the average examiner in their DDS office, only 2% (=0.23*0.08, where 0.08 is the mean absolute deviation of EXALLOW) of applicants would have a new outcome. Alternatively, if one eliminated the top and bottom 1% of DDS examiners, then marginal applicants would represent just under 11% (=0.23*(0.64-0.17)) of all applicants.

they were assigned, while 15.4% are denied for that reason. This does not imply that these decisions were necessarily incorrect, only that at least one examiner in the same DDS would have come to a different decision.

In contrast, we find that 57% of applicants would receive SSDI benefits regardless of the initial examiner assigned to their case. This is given by the estimated probability of SSDI receipt for the strictest examiner (see Table 2). Many of these "always takers" only receive benefits on appeal or reapplication. The remaining 20% of applicants are "never takers" – that is, they would never receive benefits no matter which examiner reviewed their case. Figure 5 illustrates how these estimates derive from our first stage regression model of SSDI receipt on the examiner initial allowance rate. Each point on the figure is an applicant-examiner combination, where applicants are characterized by impairment severity and examiners are characterized by their initial allowance rate. Applicant severity is on the y-axis (ordered bottom to top from most to least severe on unobserved scale u, defined below) and examiner leniency is on the x-axis (ordered left to right from most strict to most lenient). Monotonicity implies that any applicant with $u < P_A$ will receive benefits no matter which examiner is assigned to his case, and any applicant with $u > P_A + P_M$ will never receive benefits no matter which examiner is assigned to his case.[44] Only applicants for whom $P_A < u < P_A + P_M$ may or may not receive benefits depending on the examiner. It is straightforward to show that $u \equiv 1 - F(s)$ is uniformly distributed, where F is the distribution of unobserved severity s (see Section 4).[45] Since u is uniform, the percent of applicants who are always takers is P_A, the percent of marginal applicants is P_M, and the percent of never takers is $1 - P_A - P_M$. Therefore, the marginal applicants fall between the 20th and 43rd percentiles of the distribution of unobserved severity.

5.2. Characteristics of Applicants on the Margin of Program Entry

---

[44] Note that the probability of SSDI receipt for an applicant assigned to an examiner with a 0% initial allowance rate is positive because of the appeals process. Similarly, the probability of SSDI receipt for an applicant assigned to an examiner with a 100% (raw) initial allowance rate can be less than 1 after adjusting for covariates.

[45] $\Pr(u \le t) = \Pr(1 - F(s) \le t)$
$\qquad\qquad = \Pr(F(s) > 1 - t)$
$\qquad\qquad = \Pr(s > F^{-1}(1 - t))$
$\qquad\qquad = 1 - F(F^{-1}(1 - t))$
$\qquad\qquad = t$

Applicants on the margin of program entry are of significant policy interest since this group would be most affected by policy interventions affecting initial determinations, and knowledge of their characteristics could be used to improve cost forecasts of proposed policy changes. While we cannot identify which applicants are in the marginal group and which are not, our instrumental variables research strategy offers a natural opportunity to draw probabilistic inferences about their characteristics.[46] Specifically, the relative likelihood that the marginal entrant has a particular observable characteristic, compared with a typical applicant, is given by the ratio of the first stage coefficient conditional on that characteristic to the overall first stage coefficient.[47] Table 3 presents first stage estimates for groups defined by impairment type, age, and prior earnings quintile. In order to improve the precision of our estimates, we pool the 2005 and 2006 samples. The coefficient on EXALLOW is statistically significant in nearly all groups (except for miscellaneous "special/other" cases) and its magnitude varies substantially across groups. For example, the first stage coefficient for musculoskeletal cases (0.157) implies that 16 percent of musculoskeletal applicants are on the margin of program entry as a result of variation across examiners in the same DDS. On the other hand, the coefficient for mental impairments (0.322) implies that 32 percent of applicants with mental disorders are on the margin of program entry.[48]

Table 3 also presents the estimated relative likelihoods for each group. From this, we see that the marginal entrant is 43% more likely than the average applicant to suffer from a mental disorder and 30% less likely than average to suffer from a musculoskeletal impairment.[49] Similarly, the marginal entrant is 43% more likely to be either very young or relatively old (under 29 and over 60), and 22% more likely to have very low prior earnings. Thus, a policy change that uniformly increased allowance thresholds at the initial determination stage would induce entry by younger, lower-earning individuals with mental impairments.

---

[46] This insight is due to Angrist and Pischke (2009).

[47] This is a simple consequence of Bayes' rule. The first stage coefficient conditional on some characteristic X gives $\Pr(M \mid X) = P(X \mid M)P(M) / P(X)$ where M denotes marginal applicant. Dividing by the overall first stage coefficient, $P(M)$, gives the relative likelihood, $P(X \mid M) / P(X)$.

[48] Interestingly, the bulk of musculoskeletal cases are allowed at the appellate stages, while most mental impairment cases are allowed at the initial determination.

[49] Another way of characterizing the marginal population is by examining the probability that a marginal applicant has a particular characteristic X. This is just the relative likelihood times multiplied by the fraction of applicants with characteristic X. For example, though applicants with musculoskeletal impairments and mental disorders make up 40% and 21% of the applicant population, respectively, they account for roughly equal shares of the marginal population: 28% (=40*0.70) and 30% (=21*1.43), respectively.

Such compositional changes could have important implications for government spending. Individuals with mental impairments incur annual medical costs that are three times higher than medical costs for individuals with relatively mild physical impairments such as back and joint problems and cardiovascular problems (Foote and Hogan, 2001). Individuals with mental impairments also spend 50% more time on the disability insurance rolls compared to individuals with musculoskeletal impairments overall; this figure increases to more than 60% for individuals who enter between ages 18-34 (Rupp and Scott, 1996).

Finally, by computing the percent of applicants in each group who are always takers, we can benchmark differences in unobserved severity across groups of applicants. For example, only 33% of applicants from the bottom prior earnings quintile have impairments above the severity threshold of all examiners, while 74% of applicants in the top prior earnings quintile do. This pattern suggests economic conditions may motivate the application decisions of low-income workers with less severe impairments.

6. Effects of SSDI Receipt on Labor Supply
6.1. Main Estimates

In this section we present our main estimates of the average causal effect of SSDI receipt on the employment and earnings of the marginal entrant. Table 4 presents results for the labor supply regressions estimated by OLS and IV (2SLS), separately for applicants with 2005 and 2006 decisions. We examine three outcomes: employment (earning more than $1,000/year), employment above SGA (e.g., earning more than $11,760 per year in 2009) and annual earnings. We measure all labor supply outcomes two and three years after the year of the initial decision and we also examine outcomes four years later for those who received their initial determinations in 2005. In all models, we control for geography (DDS office and 3-digit zip code indicators), characteristics of the impairment (body system codes and terminal illness indicator), age, prior labor market history and seasonality (month of decision).

The OLS estimates suggest that SSDI receipt is associated with an approximately 34-35 percentage point drop in employment two years after the initial determination, a 24-25

percentage point drop in the probability of engaging in substantial gainful activity, and a roughly $7,000 drop in annual earnings.[50] Covariates have little effect on the magnitude of the estimates.

In contrast, our IV estimates imply that SSDI receipt causes a 28 percentage point decrease in employment two years after the initial determination. Thus, adjusting for differences in *unobservable* characteristics via 2SLS has a substantial impact on the estimated labor supply effects. By four years after the initial determination, the estimated employment effect is only 16 percentage points. This pattern is repeated for the other two outcomes, employment above SGA and annual earnings. Two years after the initial decision, SSDI receipt leads to a 18-19 percentage point decrease in employment above SGA and a $3,800 to $4,600 decrease in annual earnings. Collectively, our estimates imply very large labor supply effects.[51] It is worth noting, however, that the earnings estimates imply that beneficiaries would have earned only one-quarter to one-half of their pre-application earnings in the absence of benefit receipt.

Another way of obtaining the IV estimates is to estimate the reduced form effects of the examiner allowance rate on labor supply and scale by the first stage coefficient from the regression of SSDI receipt on the examiner allowance rate. Table 5 shows that a 10 percentage point increase in the examiner allowance rate decreases employment 2 years after the initial determination by 0.6-0.7 percentage points. If we attribute this difference entirely to differences in SSDI receipt, then dividing by the first stage coefficient implies that SSDI receipt decreases employment by 28 percentage points (=0.057/0.204, or 0.069/0.243, for 2005 and 2006, respectively). This assumes that the only impact examiners have on applicants is through the channel of benefit receipt.

But as Parsons (1991) noted, the SSDI application process is itself a distinct channel that may adversely affect the labor supply of applicants; in particular, denied applicants who pursue appeals spend substantially more time in process, waiting for their cases to be evaluated. While applicants await a final decision, they cannot engage in substantial gainful activity without compromising their eligibility for benefits. If time out of the labor market adversely affects

---

[50] Note that, because our measure of earnings includes zeros for nonworkers in place of *potential* earnings, the estimated earnings effects are likely understated for both OLS and IV.

[51] The computation of percent effects is not straightforward because we only observe labor supply for all allowed and all denied applicants, and not for the subset of applicants on the margin of program entry. Assuming their employment rate is between the rate for all allowed and all denied, then we find that SSDI reduces employment by 53-189%, employment above SGA by 66-384%, and earnings by 42-194%. Abadie (2003) has developed a method to estimate the counterfactual labor supply of the marginal applicant, but the method depends on having a binary instrument. Rather than discretize our instrument, we chose to bound the size of the effect.

employment outcomes (e.g., by depreciating skills, or signaling low labor force attachment to potential employers), then when denied applicants attempt to re-enter the labor market after exhausting their appeals they will have a more difficult time finding work than if they had never appealed. Thus, applicants assigned to relatively strict examiners who spend more time waiting for a decision may have lower employment rates than they otherwise would have, leading us to *underestimate* the effect of SSDI receipt on labor supply.

Figure 6 shows the relationship between our instrumental variable and average initial (DDS-level) and final processing times. Final processing times are measured as time from filing date to benefit receipt date (minus the five-month waiting period, if applicable) for SSDI beneficiaries, and time to last observed decision for non-beneficiaries.[52] While initial processing times (approximately 4 months, or 120 days on average) appear to be unrelated to the examiner allowance rates, final processing times are shorter for more lenient examiners. As a result, applicants assigned to the most lenient examiners spend approximately 125 fewer days awaiting a final allowance decision than applicants assigned to an average examiner. This pattern is the product of two forces. First, applicants denied by strict examiners are more likely to enter the appeals process, which adds on average five months at the reconsideration stage and an additional two years for applicants progressing to the ALJ level (Autor, Maestas, Mullen and Strand, 2011). On the other hand, mitigating the impact of the first channel, initially denied applicants with more severe impairments (concentrated among the stricter examiners) will spend less time in the appeals process, all else equal, since they are more likely to overturn their initial denials at earlier stages in the appeals process.

To investigate the influence of endogenous application processing time on the estimated labor supply effect, we perform a back-of-the-envelope calculation where we add an estimate of the labor supply depreciation effect to the reduced form effect of EXALLOW on labor supply in order to "restore" the lost skills due to time out of the labor market due to appeals. To do this, we use an estimate from Autor, Maestas, Mullen and Strand (2011) of the effect of application processing time on employment using exogenous variation in examiners' average initial processing times among initially allowed applicants; they estimate a depreciation effect of

---

[52] Since we do not observe higher level appeals, final processing times are censored at the ALJ level for a very small fraction of denied applicants with cases still in progress more than five years after their initial determination. It is also possible that final processing times are censored at the initial decision for denied applicants who bypass the appeals process and submit a new application.

approximately 0.4 percentage points per month, or 1.6 percentage points over 4 months (the average difference in processing time between the strictest and most lenient examiners). Adding this to the reduced form estimate and dividing by the first stage coefficient yields a modified 2SLS estimate of 35-36 percentage points for 2005 and 2006, respectively), slightly larger than the OLS estimates. This is consistent with Parson's (1991) observation that skill depreciation associated with the lengthy appeals process invalidates the interpretation of observational differences in labor supply between allowed and denied applicants as an upper bound on the effect of SSDI receipt on labor supply.

It is also of note that the magnitude of the effect of SSDI on labor supply declines over time. Certainly, SSDI work incentives become weaker with program tenure as beneficiaries are more likely to have exhausted their Trial Work Period; but this should cause the labor supply disincentive effect to *increase* over time, not decrease. Similarly, the possibility of recovery should lead to greater labor supply effects over time if many denied applicants eventually recover capacity and return to work.

One potential explanation is that as the sample ages over time, more denied applicants become eligible for and claim Social Security benefits rather than returning to work. But when we estimate our results on the subset of individuals ages 18-57 in the year of decision (i.e., no more than 61 years old four years later), the declining pattern over time still holds (see Table A-3).[53] Similarly, we rule out differential mortality as a potential explanation by estimating regressions on a balanced panel of individuals surviving at least four years after the initial decision (see Table A-4).[54]

Another explanation for the declining effect is the Great Recession. The labor supply effect of SSDI receipt reflects both the supply of *and* demand for disabled workers. A lack of employment opportunities for disabled workers in 2008 and 2009 – even for those who were willing and able to work – could lead to smaller labor supply effects in the later years of the sample period. We think the recession is unlikely to be the main cause of the declining labor supply effect for two reasons. First, the estimated effects for two years and three years later,

---

[53] Not surprisingly, the estimated labor supply effects are larger when Social Security benefit claiming is excluded as an option for denied applicants.

[54] As a final sensitivity check, we try alternative sample definitions where we include only examiners with at least 10, 100 or 150 cases (instead of 30), and we add back examiners with more than 900 cases (suspected training cases). Table A-5 shows that larger minimum values result in somewhat larger estimated labor supply effects, while adding back training cases does not affect the estimates substantially. The first stage is relatively stable across the range of cutoff values.

respectively, are similar for the 2005 and 2006 decision-year samples even though the labor supply outcomes were measured one year apart, before and after the start of the Great Recession. Second, we estimated the model stratifying by state-level unemployment rates in 2009 (not shown) and did not find any discernable pattern across states affected differentially by the Great Recession.

The declining pattern in labor supply effects could also reflect the confounding effect of long processing times for those who appeal. On the one hand, if more and more denied applicants return to the labor market over time as they complete appeals, then the labor supply disincentive effect should actually increase. But if the confounding depreciation effect of processing time increases over time as final processing times unfold, then the estimated labor supply effect could fall even if the true disincentive effect remains constant.

Finally, the observed pattern is consistent with denied applicants perhaps trying but *failing* to sustain employment. This would imply that the work capacity of SSDI beneficiaries on the margin of program entry diminishes with time since onset of the disability.

6.2. Heterogeneity in the Effect of SSDI Receipt on Labor Supply

Our main estimates imply that the employment rate of new beneficiaries would have been 28 percentage points higher in the absence of benefit receipt. However, there is no reason to expect this effect to be the same for all beneficiaries; those with more severe impairments should have less employment capacity than those with less severe impairments. If this is the case, the effect of SSDI receipt on labor supply should vary with impairment severity. Our unique research strategy, leveraging variation in allowance thresholds across thousands of examiners, offers a first opportunity to test for heterogeneity in the effect of SSDI receipt on labor supply In this section, we show that the effect of SSDI receipt varies along both observable and unobservable dimensions. This implies that the 2SLS estimates from the previous section recover a local average treatment effect (LATE), where the causal effect of SSDI receipt is averaged over the population of applicants on the margin of program entry.

In Table 6 we present estimates of the effect of SSDI receipt on employment two years after the initial decision for groups defined by impairment type, age, and prior earnings quintile. The differences across impairment types are striking. SSDI receipt causes a 17.3 percentage point reduction in employment among applicants with musculoskeletal impairments compared to

a 36.6 percentage point reduction for applicants with mental disorders—more than twice as large.[55] The effects among those with malignant neoplastic diseases (i.e., cancer) and cardiovascular, neurological, and digestive disorders are of similar magnitude as the effect for mental disorders. We estimate even larger labor supply effects for applicants with endocrine system diseases (e.g., diabetes) and skin disorders (49.4 and 91.6 percentage points, respectively).[56] We estimate small and statistically insignificant labor supply effects among those with respiratory, immune system, special senses and speech, genitourinary, and hematological disorders.

We also test for heterogeneous labor supply effects by age and by prior earnings quintile. Using observational comparisons of allowed and denied applicants, von Wachter et al. (2012) found larger differences in labor supply among younger applicants (ages 30-44) than older applicants (45-64). This pattern holds for the causal effect as well. Our IV estimates imply that SSDI receipt causes a 34.2 percentage point reduction in employment among younger applicants compared to a 26.6 percentage point reduction among older applicants. The patterns by income quintile point to smaller employment effects for applicants in the top prior earnings quintile, consistent with our earlier finding that applicants in the top quintile tend to have more severe impairments (see Section 5.2).

More generally, the differences in labor supply effects across groups could be due to differences in unobservable characteristics such as impairment severity. Because applicants are (conditionally) randomly assigned to a large number of examiners, each with a different allowance threshold, we can trace out the effect of SSDI receipt along different margins of entry by estimating marginal treatment effects (MTE) (Heckman, Urzua, Vytlacil, 2006). As we showed in Section 5.1, under monotonicity these margins correspond to percentiles of the distribution of unobserved severity. This exercise is particularly interesting in that it tests for heterogeneous treatment effects on the basis of *unobservables*.

Practically, computation of the MTE can be accomplished by computing numerical derivatives of a smoothed function relating mean labor supply outcome, $E[y]$, to the predicted

---

[55] This may partly reflect differences in decision times across the two groups. Recall that applicants with musculoskeletal impairments were much more likely than applicants with mental impairments to appeal. Indeed, we estimate similar fractions of always takers within the two groups (see Table 3).

[56] While we report the IV estimate for "special/other" impairments for completeness, there is no first stage for this group (see Table 3) and as a result the second stage estimate is not meaningful.

probability of SSDI receipt. Specifically, we regress initial allowance decisions on indicators for type of impairment, age group, decision month and DDS, as well as a measure of average prior earnings, and construct the residual, Z, which by construction is orthogonal to the case mix controls and varies systematically only with EXALLOW. Then we estimate a probit of ultimate SSDI receipt on the residualized Z. This is our measure of the predicted probability of SSDI receipt, $P(Z)$. Next we estimate a local quadratic regression of employment on predicted SSDI receipt and compute the numerical derivative of this function to estimate $\partial E[y]/\partial P(Z)$.

Figure 7 shows the MTE as a function of unobserved severity, where severity is reverse ordered and measured in percentiles (see definition of u in Section 5.1), along with boot-strapped 95% confidence intervals. Applicants on the margin for an examiner with a predicted SSDI receipt rate of 65% (the mean rate) are in the 65th percentile of the unobserved (reverse) severity distribution. That is, they have an impairment that is less severe than 65% of applicants, and more severe than 35% of applicants. Since we estimate that 57% of applicants are always-takers (that is, they would receive SSDI benefits regardless of initial examiner assignment), the MTE is not identified for applicants on the margin of SSDI receipt rates less than 57%. Similarly, the MTE is not identified for applicants on the margin of SSDI receipt rates greater than 80% (=57+23, the fraction of marginal applicants). As a result, we are only able to trace the MTE for applicants between the 57th and 80th percentiles of the unobserved (reverse) severity distribution (or the 20th to 43rd percentiles of the actual unobserved severity distribution s). The estimates become imprecise at the more extreme ends of the distribution since there are relatively small numbers of examiners with margins at these points.

The estimated MTE declines monotonically as unobserved severity falls (or reverse severity rises); because it is negative in value, as it declines it increases in magnitude, implying the disincentive effect of SSDI rises as the allowance threshold is lowered and applicants with less severe impairments are allowed on the program. The MTE is not statistically different from zero among applicants with the highest severity impairments in the marginal population, but increases in magnitude to more than 50 percentage points among those in the marginal population with the least severe impairments.[57] This implies that beneficiaries with relatively

---

[57] Recalling that average processing times fall with examiner allowance rates above the average, the MTE may increase in magnitude at an even *greater* rate in this part of the distribution if longer processing times affect employment independently of SSDI receipt.

27

severe impairments have little work capacity, while those with the least severe impairments have a very large degree of work capacity.

Our finding that the MTE is heterogeneous in unobserved severity is important for interpreting our 2SLS estimate as a LATE and comparing our estimate to other estimates from the literature. Of particular interest is the comparison of our estimate of 28 percentage points with French and Song's (2011) estimate of 26 percentage points based on applicants on the margin of allowance at the ALJ level among those who appeal their initial determination. In light of the interpretation of both estimates as LATEs, it should not be surprising that French and Song is similar to ours. In both cases, the estimated effects are averaged among the *marginal* program entrants—that is, excluding the always takers and never takers. Although the French-Song sample excludes initially allowed applicants, as well as initially denied applicants who do not pursue their cases further, these groups are included among the always takers and never takers, respectively, in our framework and thus they do not contribute to the estimation. By the same token, the French-Song sample also includes always takers (those who win their appeal no matter to which judge they are assigned) and never takers (those who lose their appeal no matter to which judge they are assigned), and these are also always takers and never takers, respectively, in our framework. Thus, the margin of allowance is the same margin in both natural experiments and it makes sense that both estimates are similar.

7. Conclusion

This paper presents causal estimates of the disincentive effect of the SSDI program on the labor supply of program entrants. Ours is the first set of estimates of the disincentive effect estimated using the entire population of SSDI applicants in the present day system. In addition, our quasi-experimental research design applied to a new administrative data set facilitates examination of the important heretofore unanswered policy questions of whether the program disincentive effect varies across individuals, to what extent, and in what ways. In particular, we can assess the extent to which the disincentive effect varies with unobservable impairment severity. This is of particular interest since over the last two decades, the SSDI caseload has become increasingly dominated by individuals with impairments that are particularly difficult to assess, such as mental and musculoskeletal impairments.

We find that employment of the marginal program entrant would be on average 28 percentage points greater in the absence of SSDI benefit receipt two years after the initial determination. His or her likelihood of engaging in substantial gainful activity as defined by the SSDI program would be on average 18-19 percentage points higher, and he or she would earn $3,800 to $4,600 more per year on average in the absence of SSDI benefit receipt. We estimate that in the present system, 23% of applicants are on the margin of program entry, that is their ultimate outcome depends on their initial examiner assignment. The marginal SSDI entrant is more likely to have a mental disorder, be younger, and have pre-onset earnings in the lowest earnings quintile. Such individuals tend to have higher expected medical costs and greater expected program duration.

Importantly, we also find that the SSDI labor supply disincentive effect is not constant across individuals. Over a range corresponding to approximately 10 percentage points above and below the current system allowance rate, the marginal treatment effect varies from being statistically indistinguishable from zero for individuals characterized by high unobservable severity, to more than 50 percentage point reduction in employment for those with lower unobservable severity. Our estimates point to markedly greater residual work capacity among beneficiaries with (unobservably) less severe impairments, and imply that a policy with the effect (intended or unintended) of relaxing access to SSDI benefits (regardless of impairment type) in the initial determination phase would lead to an increase in the program's labor supply disincentive effect.

Finally, we close with two important qualifications. First, our estimates of the employment potential of beneficiaries had they not received SSDI benefits reflect the population of applicants, examiners and the program rules in force in 2005 and 2006, and the labor market conditions in the years following. Subsequent changes in economic conditions, disability policy, or healthcare policy could alter the composition of applicants and beneficiaries in different ways, creating different labor supply disincentives (or incentives) at that time. Second, we have presented estimates of the effect of SSDI *benefit receipt*, and therefore our estimates are not a measure of labor supply in the absence of the program itself. If the SSDI program diminishes labor supply through channels other than benefit receipt, such as the application and appeals process, then our estimate of the effect of benefit receipt is just one component of the program's total effect on labor supply.

References

Abadie, Alberto (2003). "Semiparametric instrumental variable estimation of treatment response models." *Journal of Econometrics*, 113: 231-263.

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.

Angrist, Joshua D. and Jorn-Steffen Pischke (2009). Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press: Princeton, New Jersey.

Autor, David, and Mark Duggan (2003). "The Rise in the Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics*, 118(1): 157–206.

Autor, David, Frank Levy, and Richard Murnane (2003). "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics*, 118(4): 1279-1333.

Autor, David, Nicole Maestas, Kathleen Mullen and Alexander Strand (2011). "What are the Real Application Costs of SSDI? The Effect of Waiting Time on Labor Force Participation and Earnings." Michigan Retirement Research Center Research Paper No. 2011-258.

Board of Trustees of the Federal Old Age and Survivors Insurance and Disability Insurance Trust Funds (2012). "The 2011 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds." Accessed on 6-8-2012 at: http://www.socialsecurity.gov/OACT/TR/2012/tr2012.pdf.

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

Bound, John, and Timothy Waidmann (1992). "Disability Transfers, Self-Reported Health, and the Labor Force Attachment of Older Men: Evidence from the Historical Record." *Quarterly Journal of Economics*, 107(4): 1393–419.

Burkhauser, Richard, and Mary Daly (2011). *The Declining Work and Welfare of People with Disabilities: What Went Wrong and a Strategy for Change.* Washington, DC: AEI Press.

Chen, Susan and Wilbert van der Klaauw (2008). "The Work Disincentive Effects of the Disability Insurance Program in the 1990s." *Journal of Econometrics*, 142: 757-784.

Christensen, Kaare, G. Doblhammer, Roland Rau, and James W. Vaupel (2009). "Ageing Populations: The Challenges Ahead." *Lancet*, 374:1196-208.

Doyle, Joseph (2007). "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review*, 97(5):1583-1610.

Doyle, Joseph (2008). "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care." *Journal of Political Economy*, 116(4): 746-770.

Duggan, Mark, and Scott Imberman (2009). "Why Are the Disability Rolls Skyrocketing? The Contribution of Population Characteristics, Economic Conditions, and Program Generosity." *Health at Older Ages: The Causes and Consequences of Declining Disability among the Elderly*. David Cutler and David Wise, editors, pp. 337-379. Cambridge, MA: National Bureau of Economic Research.

Foote, S. M. and C. Hogan (2001). "Disability Profile And Health Care Costs Of Medicare Beneficiaries Under Age Sixty-Five." *Health Affairs*, 20(6): 242-253.

French, Eric and Jae Song (2011). "The Effect of Disability Insurance Receipt on Labor Supply." Federal Reserve Bank of Chicago Working Paper Series, WP 2009-05.

French, Eric and Christopher Taber (2010). "Identification of Models of the Labor Market," in *Handbook of Labor Economics*, Vol. 4a, pp. 537-617.

Gallicchio, Sal and Barry Bye (1981). "Consistency of Initial Disability Decisions Among & Within States." U.S. Department of Health and Human Services, Social Security Administration, Office of Policy, Office of Research and Statistics.

Government Accountability Office (GAO) (2004). "Social Security Administration: More Effect Needed to Assess Consistency of Disability Decisions." Report to the Chairman, Subcommittee on Social Security, Committee on Ways and Means, House of Representatives, GAO-04-656.

Gruber, Jonathan and Jeffrey Kubik (1997). "Disability Insurance Rejection Rates and the Labor Supply of Older Workers." *Journal of Public Economics*, 64: 1-23.

Gruber, Jonathan (2000). "Disability Insurance Benefits and Labor Supply." *Journal of Political Economy*, 108(6): 1162-1183.

Haveman, Robert, and Barbara Wolfe (1984a). "Disability Transfers and Early Retirement: A Causal Relationship?" *Journal of Public Economics*, 24(1): 47–66.

Haveman, Robert, and Barbara Wolfe. (1984b). "The Decline in Male Labor Force Participation: Comment." *Journal of Political Economy*, 92(3): 532–41.

Heckman, James J., Sergio Urzua, and Edward Vytlacil (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economics and Statistics*, 88(3): 389–432.

Houtenville, Andrew, David Stapleton, Robert Weathers II, and Richard Burkhauser, editors (2009). *Counting Working-age People with Disabilities: What Current Data Tell Us and Options for Improvement*. Kalamazoo, MI: W.E. UpJohn Institute for Employment Research.

Kane, Thomas J. and Douglas O. Staiger (2008). "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper No. 14607.

Kling, Jeffrey R. (2006). "Incarceration Length, Employment and Earnings." *American Economic Review*, 96(3): 863-876.

Lewin Group, Inc. (2001). "Evaluation of SSA's Disability Quality Assurance (QA) Processes and Development of QA Options That Will Support the Long-Term Management of the Disability Program." Final Report submitted to the Social Security Administration under Contract Number 0600-96-27331, March 16, 2001.

Livermore, Gina, David Stapleton and Henry Claypool (2009). "Health Insurance and Health Care Access Before and After SSDI Entry." The Commonwealth Fund pub. No. 1255.

Nagi, Saad (1969). Disability and Rehabilitation: Legal, Clinical, and Self-Concepts and Measurement. Columbus, OH: Ohio State University Press.

Office of the Inspector General, Social Security Administration (2006). "The Social Security Administration's Identification of Special Disability Workload Cases." Evaluation Report A-13-05-15028.

Office of the Inspector General, Social Security Administration (2008). "Disability Claims Overall Processing Times." Audit Report A-01-08-18011.

Parsons, Donald O. (1980). "The Decline of Male Labor Force Participation." *Journal of Political Economy*, 88(1): 117-34.

Parsons, Donald O. (1982). "The Male Labour Force Participation Decision: Health, Reported Health, and Economic Incentives." *Economica*, 49(193): 81–91.

Parsons, Donald O. (1991). "The Health and Earnings of Rejected Disability Insurance Applicants: Comment." *American Economic Review*, 81(5): 1419-1426.

Perry, Cynthia (2008). "Does Treating Maternal Depression Improve Child Health Management? The Case of Pediatric Asthma." *Journal of Health Economics*, 27: 157-173.

Rupp, Kalman and Charles G. Scott (1996). "Trends in the Characteristics of DI and SSI Disability Awardees and Duration of Program Participation." *Social Security Bulletin*, 59(1):3-21.

Social Security Administration (2010). "Annual Statistical Report on the Social Security Disability Insurance Program, 2009." Available online at http://www.ssa.gov/policy/docs/statcomps/di_asr/2009/index.html.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo (2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business and Economic Statistics*, 20(4):518-529.

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.

Table 1. Summary Statistics

Variable	All Applicants	Initially Allowed	Initially Denied		
			Allowed on Appeal	Denied on Appeal	No Appeal
No. observations	2,214,066	723,627	719,705	243,559	527,175
Percent of sample	100.0%	32.7%	32.5%	11.0%	23.8%
Percent of initial denials	--	--	48.3%	16.3%	35.4%
Initial decision (final step)					
2. Denied - nonsevere	17.3%	0.0%	17.8%	28.5%	35.0%
3. Allowed - met listings	13.8%	42.3%	0.0%	0.0%	0.0%
4. Denied - work past jobs	24.0%	0.0%	38.6%	33.0%	33.0%
5. Denied - able to work	26.0%	0.0%	43.7%	38.5%	31.9%
5. Allowed - unable to work	18.9%	57.7%	0.0%	0.0%	0.0%
Body system code					
Musculoskeletal system	37.9%	27.0%	45.6%	42.5%	40.3%
Mental disorders	20.6%	25.7%	16.3%	20.5%	19.4%
Cardiovascular system	8.4%	9.2%	8.9%	6.7%	7.2%
Neurological	8.2%	11.3%	7.1%	6.2%	6.2%
Endocrine system	4.3%	2.6%	5.4%	4.9%	4.6%
Respiratory system	4.0%	5.4%	3.4%	3.5%	3.2%
Special/other	3.9%	1.0%	3.3%	6.6%	7.3%
Malignant neoplastic diseases	3.5%	6.2%	2.1%	1.2%	2.7%
Immune system disorders	2.7%	3.1%	2.7%	2.5%	2.2%
Digestive system	2.4%	1.9%	2.6%	2.5%	2.8%
Special senses and speech	2.2%	2.7%	1.5%	1.8%	2.5%
Genitourinary impairments	1.4%	3.1%	0.6%	0.5%	0.7%
Skin disorders	0.3%	0.3%	0.3%	0.3%	0.4%
Hematological disorders	0.3%	0.3%	0.3%	0.3%	0.4%
Terminal illness case	0.8%	2.1%	0.2%	0.1%	0.2%
Age at initial decision					
Mean	47.09	51.03	47.18	42.97	43.46
Std. deviation	10.36	10.49	9.20	9.92	11.98
Avg. earnings, 3-5 years before					
Mean	22,697	29,123	23,358	14,495	16,764
Std. deviation	27,298	33,971	24,494	20,422	25,145
EXALLOW					
Mean	0.36	0.39	0.34	0.34	0.35
Std. deviation	0.10	0.11	0.10	0.09	0.10

Sources: DIODS, initial applications decided in 2005-2006. CPMS, ALJ-level appeals though June 2011. MBR, allowances on appeal at higher levels or additional application through December 2010. EXALLOW is examiner initial allowance rate less own decision.

Table 2. First Stage Regressions: Effect of EXALLOW on SSDI Receipt

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2005							
Coeff. on EXALLOW	0.278***	0.271***	0.220***	0.215***	0.209***	0.204***	0.204***
t-stat	31.01	30.45	26.56	25.97	26.82	26.15	26.1
Fraction of always-takers (P_A)							0.59***
t-stat							3.53
R-squared	0.012	0.017	0.042	0.043	0.118	0.128	0.128
2006							
Coeff. on EXALLOW	0.309***	0.306***	0.258***	0.254***	0.245***	0.242***	0.243***
t-stat	31.73	31.66	30.88	30.7	32.48	32.16	32.28
Fraction of always-takers (P_A)							0.56***
t-stat							3.32
R-squared	0.013	0.018	0.044	0.045	0.117	0.127	0.128
2005 and 2006							
Coeff. on EXALLOW	0.294***	0.289***	0.240***	0.235***	0.227***	0.224***	0.226***
t-stat	41.87	40.73	38.81	38.09	40.2	39.33	39.9
Fraction of always-takers (P_A)							0.57***
t-stat							3.42
R-squared	0.012	0.017	0.042	0.043	0.117	0.127	0.127
Control variables included							
3-digit zipcode		X	X	X	X	X	X
Body system codes			X	X	X	X	X
Terminal Illness indicator				X	X	X	X
Age group dummies					X	X	X
Avg. previous earnings						X	X
Month dummies							X
No. control variables	112	1015	1031	1031	1040	1041	1052

Notes: standard errors clustered at examiner level; t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.10. Mean of dependent (independent) variable = 0.65 (0.37). All regressions include DDS dummies. Pooled specification includes year interacted with month dummies. Number of observations is 1,090,345 for 2005 and 1,123,721 for 2006.

Table 3. Heterogeneity: First Stage Regressions, 2005 and 2006 Combined

	No. obs.	Initial Allowance Rate	Ultimate Allowance Rate	Fraction of always-takers (P_A)	Coeff. on EX-ALLOW	t-Stat.	Relative Likeli-hood
Body system code							
Musculoskeletal system	839,847	23.3%	62.4%	0.567	0.157	15.51	0.70
Mental disorders	455,433	40.8%	66.6%	0.547	0.322	25.24	1.43
Cardiovascular system	185,063	36.1%	70.6%	0.617	0.252	14.78	1.11
Neurological	181,162	45.3%	73.5%	0.636	0.274	17.74	1.21
Endocrine system	94,156	20.0%	61.4%	0.579	0.098	3.69	0.43
Respiratory system	88,578	43.7%	71.3%	0.656	0.162	7.30	0.72
Special/other	85,587	8.7%	36.3%	0.370	-0.019	-0.66	-0.08
Malignant neoplastic diseases	77,021	58.2%	77.9%	0.667	0.297	16.34	1.31
Immune system disorders	59,188	37.5%	70.0%	0.570	0.348	13.41	1.54
Digestive system	53,224	26.4%	60.9%	0.516	0.260	8.49	1.15
Special senses and speech	48,396	41.1%	63.7%	0.550	0.240	7.32	1.06
Genitourinary impairments	31,837	71.1%	84.2%	0.742	0.264	11.78	1.17
Skin disorders	7,483	28.2%	57.7%	0.458	0.339	3.55	1.50
Hematological disorders	7,091	32.8%	62.5%	0.500	0.344	3.77	1.52
Age at decision							
18-29	193,824	22.1%	41.7%	0.301	0.324	18.42	1.43
30-39	342,535	20.3%	50.6%	0.422	0.235	15.97	1.04
40-49	627,727	20.8%	60.8%	0.541	0.185	18.31	0.82
50-59	772,263	41.6%	78.2%	0.715	0.186	25.37	0.82
60-64	277,717	57.3%	73.3%	0.615	0.320	25.01	1.42
30-44	621,664	20.2%	53.5%	0.460	0.211	19.24	0.93
45-64	1,398,578	39.7%	73.6%	0.657	0.217	34.81	0.96
Avg. prior earnings							
Bottom quintile	442,814	19.7%	42.6%	0.328	0.275	21.80	1.22
2nd quintile	442,813	26.3%	60.3%	0.517	0.243	21.77	1.08
3rd quintile	442,813	31.3%	68.3%	0.605	0.220	20.85	0.97
4th quintile	442,813	38.2%	74.5%	0.668	0.214	21.90	0.95
Top quintile	442,813	47.9%	80.2%	0.735	0.178	20.98	0.79

Notes: standard errors clustered at examiner level; *** p<0.01, ** p<0.05, * p<0.10. Control variables include: DDS dummies, 3-digit zipcodes, body system codes, terminal illness indicator, age group dummies, avg. previous earnings, and month dummies.

Table 4. Effects of SSDI Receipt on Employment and Earnings

Outcome	2005 Decisions						2006 Decisions			
	2 Years After Decision		3 Years After Decision		4 Years After Decision		2 Years After Decision		3 Years After Decision	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
1) Earn >=$1,000/year										
Mean dependent variable | allowed	0.148		0.128		0.106		0.133		0.103	
Mean dependent variable | denied	0.522		0.515		0.471		0.489		0.448	
Coeff. on ALLOW	-0.347***	-0.279***	-0.361***	-0.227***	-0.345***	-0.158***	-0.335***	-0.283***	-0.330***	-0.207***
	(-322.48)	(-8.64)	(-336.60)	(-6.99)	(-321.81)	(-4.83)	(-297.39)	(-11.25)	(-285.84)	(-8.51)
R-squared	0.200	0.195	0.218	0.200	0.209	0.171	0.194	0.192	0.195	0.177
2) Earn >= SGA										
Mean dependent variable | allowed	0.050		0.043		0.033		0.042		0.030	
Mean dependent variable | denied	0.293		0.302		0.270		0.267		0.242	
Coeff. on ALLOW	-0.242***	-0.192***	-0.255***	-0.166***	-0.233***	-0.113***	-0.226***	-0.175***	-0.213***	-0.152***
	(-256.29)	(-7.62)	(-264.05)	(-6.70)	(-252.71)	(-4.59)	(-218.54)	(-8.82)	(-208.98)	(-8.16)
R-squared	0.149	0.144	0.166	0.152	0.156	0.128	0.141	0.136	0.140	0.132
3) Earnings										
Mean dependent variable | allowed	1,951		1,737		1,494		1,688		1,365	
Mean dependent variable | denied	8,928		9,191		8,496		8,044		7,562	
Coeff. on ALLOW	-7435***	-3781***	-7715***	-3007***	-7221***	-1,716	-6915***	-4581***	-6580***	-3308***
	(-126.50)	(-3.05)	(-182.51)	(-2.92)	(-176.24)	(-1.60)	(-84.14)	(-5.93)	(-117.85)	(-4.60)
R-squared	0.133	0.117	0.145	0.114	0.125	0.084	0.139	0.131	0.129	0.112
No. observations	1,090,345		1,069,494		1,042,666		1,123,721		1,094,347	

Notes: standard errors clustered at examiner level; t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.10. ALLOW denotes actual award decision (rather than initial decision). Control variables include: DDS dummies, 3-digit zipcodes, body system codes, terminal illness indicator, age group dummies, avg. previous earnings, and month dummies.

Table 5. Reduced Form Regressions: Effect of EXALLOW on Employment and Earnings

| Outcome	2005 Decisions			2006 Decisions	
	2 Years	3 Years	4 Years	2 Years	3 Years
1) Earn >=$1,000/year					
Coeff. on EXALLOW	-0.057***	-0.047***	-0.033***	-0.069***	-0.050***
	(-8.05)	(-6.60)	(-4.55)	(-10.51)	(-8.00)
R-squared	0.082	0.086	0.077	0.077	0.070
2) Earn >= SGA					
Coeff. on EXALLOW	-0.039***	-0.034***	-0.023***	-0.042***	-0.037***
	(-7.21)	(-6.23)	(-4.32)	(-8.34)	(-7.79)
R-squared	0.049	0.054	0.051	0.046	0.044
3) Earnings					
Coeff. on EXALLOW	-772***	-619***	-355	-1,114***	-804***
	(-2.97)	(-2.83)	(-1.57)	(-5.77)	(-4.49)
R-squared	0.066	0.062	0.054	0.071	0.058
No. observations	1,090,345	1,069,494	1,042,666	1,123,721	1,094,347

Notes: standard errors clustered at examiner level; t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.10. Control variables include: DDS dummies, 3-digit zipcodes, body system codes, terminal illness indicator, age group dummies, avg. previous earnings, and month dummies.

Table 6. Heterogeneity: Effects of SSDI Receipt on Employment (2 Years Later), 2005 and 2006 Combined

|	No. obs.	Mean Empl. | Allowed	Mean Empl. | Denied	OLS Coeff. On ALLOW	t-Stat.	IV Coeff. On ALLOW	t-Stat.
Body system code							
Musculoskeletal system	839,847	0.125	0.520	-0.361	-300.48	-0.173	-3.42
Mental disorders	455,433	0.171	0.535	-0.333	-210.25	-0.366	-11.73
Cardiovascular system	185,063	0.116	0.462	-0.328	-134.18	-0.335	-5.81
Neurological	181,162	0.119	0.538	-0.386	-144.34	-0.359	-7.40
Endocrine system	94,156	0.130	0.467	-0.316	-100.41	-0.494	-2.14
Respiratory system	88,578	0.099	0.462	-0.309	-84.61	-0.061	-0.51
Special/other	85,587	0.132	0.322	-0.222	-57.78	2.000	0.59
Malignant neoplastic diseases	77,021	0.212	0.621	-0.390	-94.75	-0.347	-6.18
Immune system disorders	59,188	0.170	0.531	-0.330	-65.02	-0.048	-0.61
Digestive system	53,224	0.158	0.523	-0.353	-80.64	-0.324	-3.08
Special senses and speech	48,396	0.158	0.529	-0.344	-77.59	-0.158	-1.37
Genitourinary impairments	31,837	0.185	0.540	-0.327	-41.38	-0.084	-0.74
Skin disorders	7,483	0.156	0.560	-0.377	-32.25	-0.916	-3.66
Hematological disorders	7,091	0.219	0.590	-0.336	-26.35	0.116	0.44
Age at decision							
18-29	193,824	0.289	0.645	-0.364	-154.42	-0.336	-6.24
30-39	342,535	0.201	0.581	-0.397	-184.06	-0.369	-7.09
40-49	627,727	0.153	0.516	-0.379	-266.59	-0.355	-7.56
50-59	772,263	0.111	0.443	-0.348	-234.89	-0.290	-8.42
60-64	277,717	0.093	0.223	-0.140	-79.17	-0.105	-3.47
30-44	621,664	0.183	0.561	-0.395	-273.58	-0.342	-8.06
45-64	1,398,578	0.115	0.418	-0.317	-312.78	-0.266	-11.13
Avg. prior earnings							
Bottom quintile	442,814	0.137	0.336	-0.202	-155.37	-0.296	-7.81
2nd quintile	442,813	0.153	0.542	-0.357	-241.28	-0.319	-7.69
3rd quintile	442,813	0.141	0.598	-0.427	-271.37	-0.311	-7.30
4th quintile	442,813	0.132	0.625	-0.466	-280.66	-0.305	-7.34
Top quintile	442,813	0.140	0.616	-0.449	-250.98	-0.192	-3.76

Notes: standard errors clustered at examiner level; *** p<0.01, ** p<0.05, * p<0.10. Control variables include: DDS dummies, 3-digit zipcodes, body system codes, terminal illness indicator, age group dummies, avg. previous earnings, and month dummies.

Figure 1. The Rise of SSDI and the Decline in Employment of the Disabled
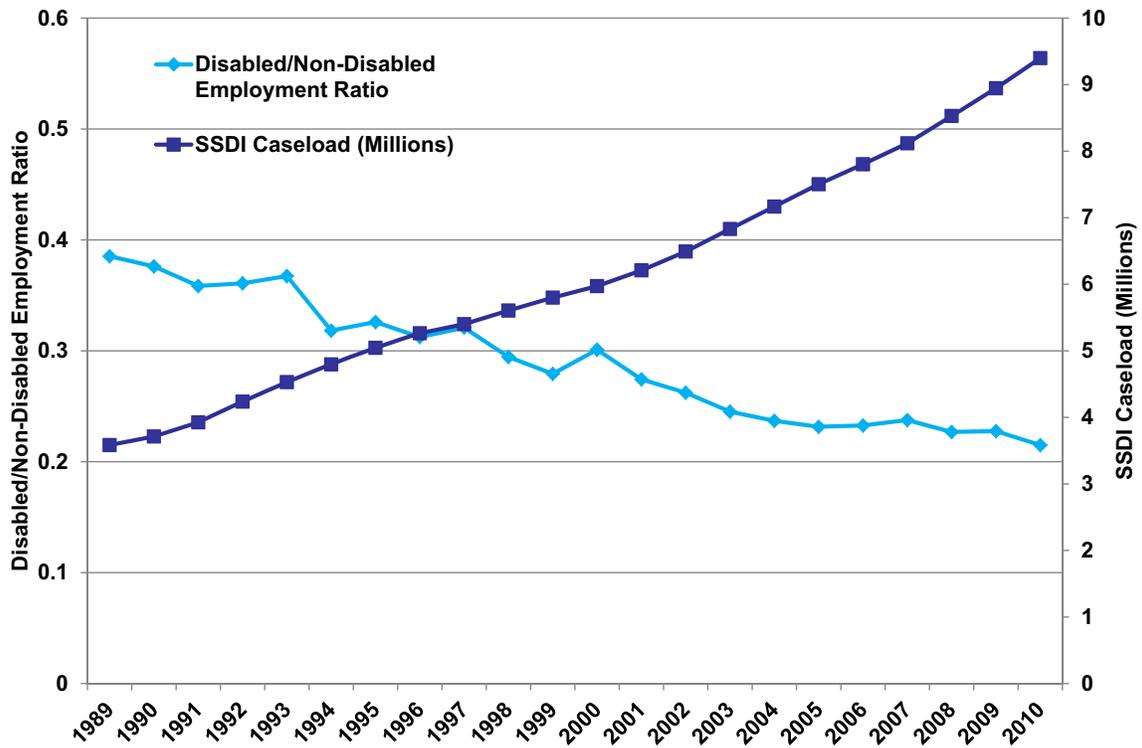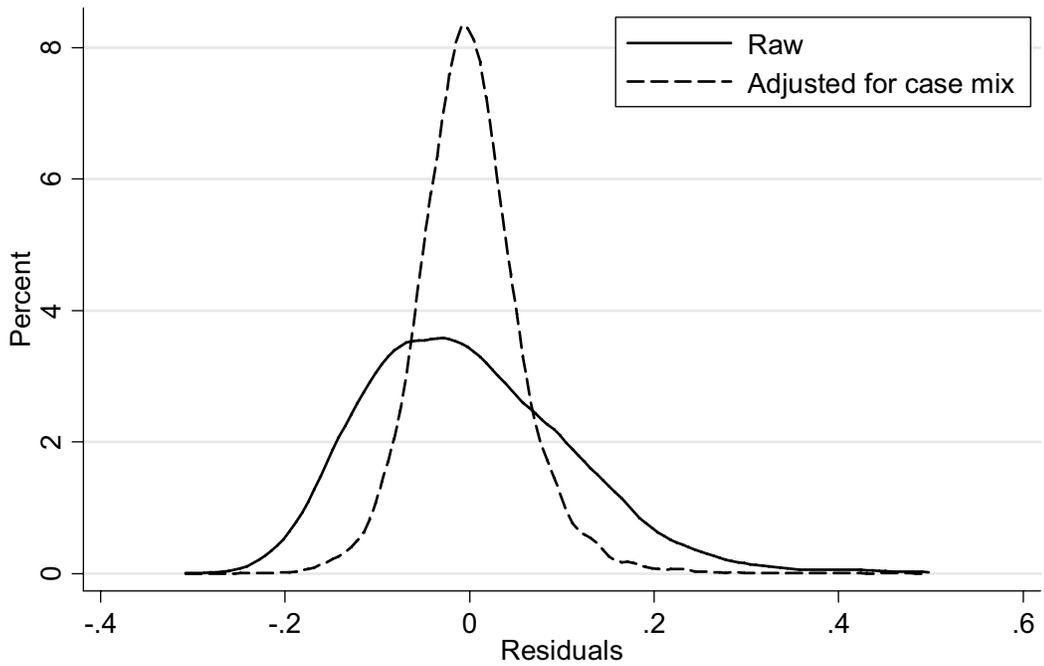

Source: Authors' tabulations of Current Population Survey; Annual Statistical Report on the Social
Security Disability Insurance Program, 2010, Table 3.

## Figure 2. Employment Before and After Initial Decision
2005 Decisions
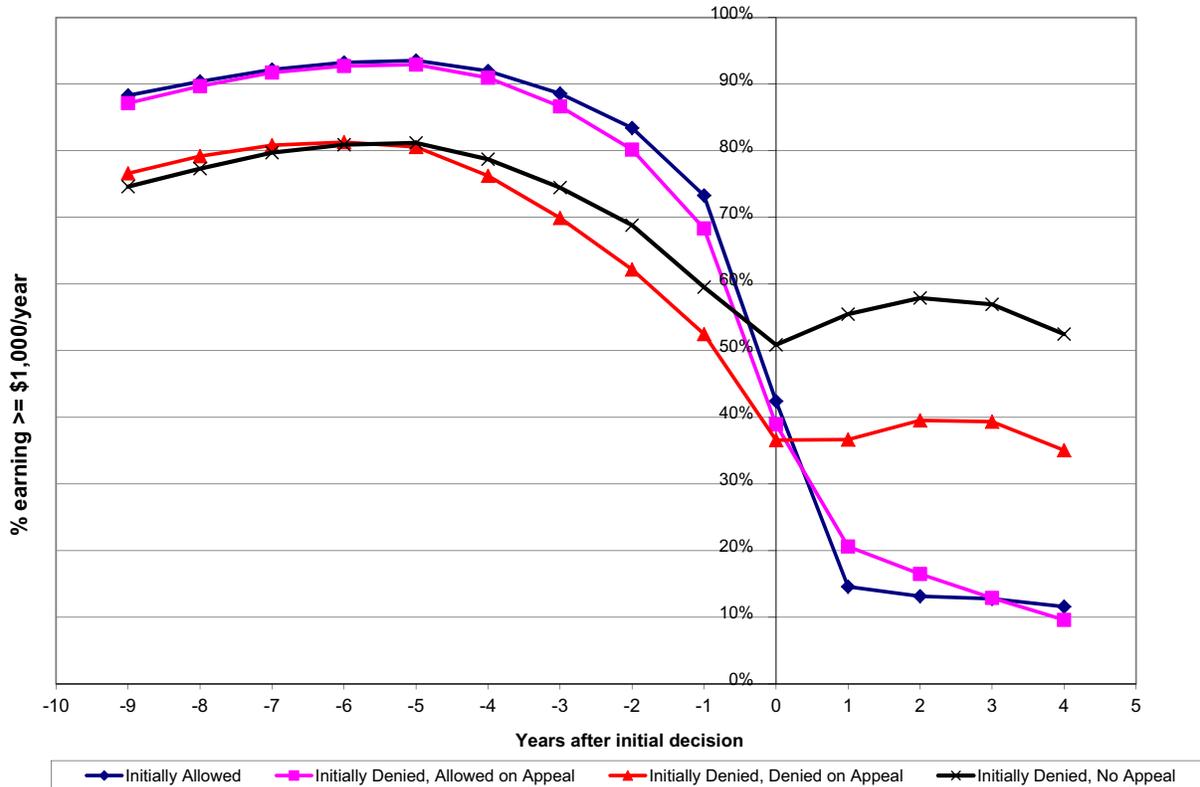
Figure 3. Distribution of Examiner Deviations
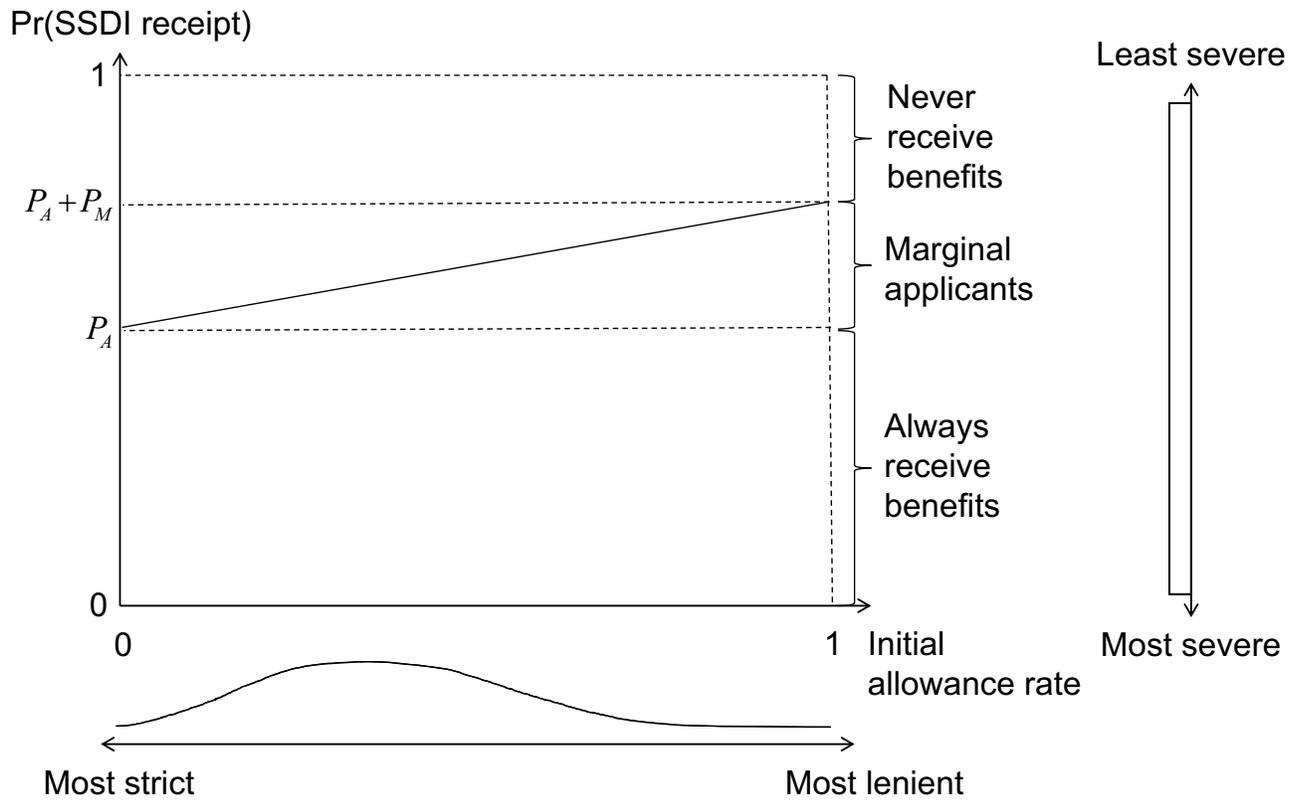## from DDS Mean Initial Allowance Rate

Source: 2005 - 2006 DIODS Data.
Caseload characteristics include DDS office, age, pre-onset earnings, body code,
3-digit ZIP code, terminal illness diagnosis and decision month.

## Figure 4. SSDI Receipt and Labor Supply
by Initial Allowance Rate
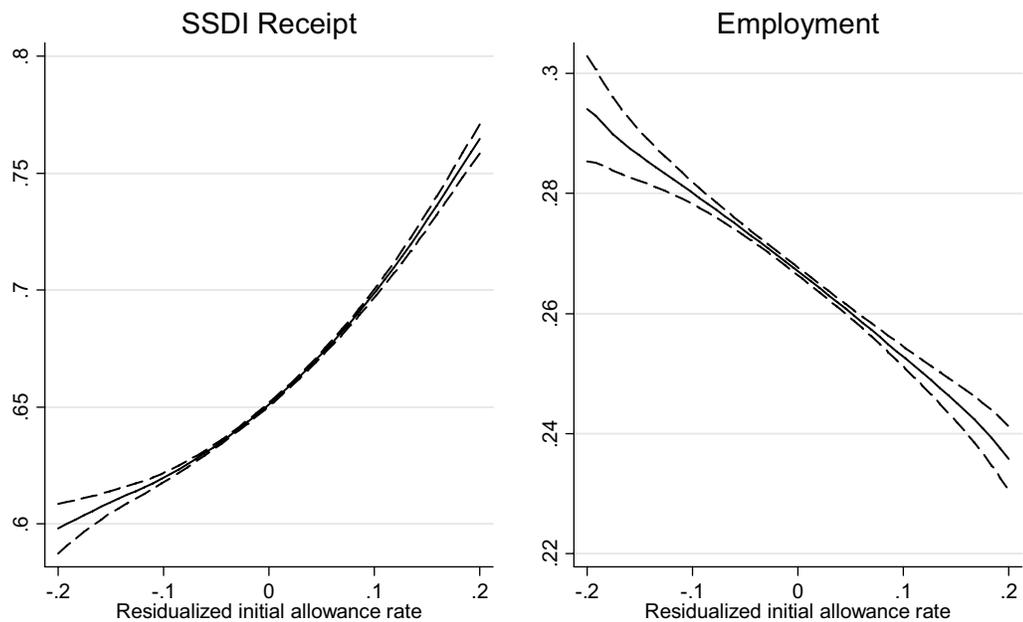
Source: DIODS data for 2005 and 2006.
95% confidence intervals shown with dashed lines.
Employment measured in the second year after the initial decision.
Bandwidth is 0.116 for DI and 0.130 for labor force participation.

Figure 5. Model of Effect of Initial Allowance Rate on Probability of SSDI Receipt

Pr(SSDI receipt)

Least severe

1

$P_A + P_M$

P_A

Never receive benefits

Marginal applicants

Always receive benefits

0

0

1

Initial allowance rate

Most severe

Most strict

Most lenient

**Figure 6. Initial and Final Processing Times
by Initial Allowance Rate**

Days

500

400

300

200

100

-.2

-.1

0

.1

.2

Residualized initial allowance rate

Initial decision time   Final decision time
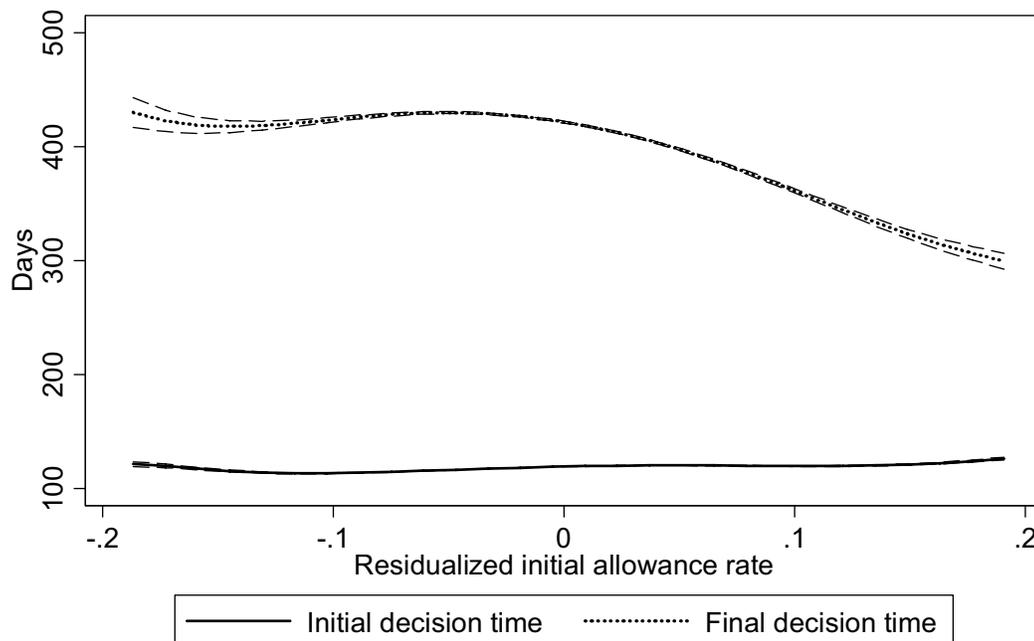
Source: DIODS data for 2005 and 2006.
95% confidence intervals shown with dashed lines.
Final processing times are measured through the appeals stage.
Bandwidth is 0.067 for initial and 0.082 for final processing time.

Figure 7. Marginal Treatment Effect on Employment
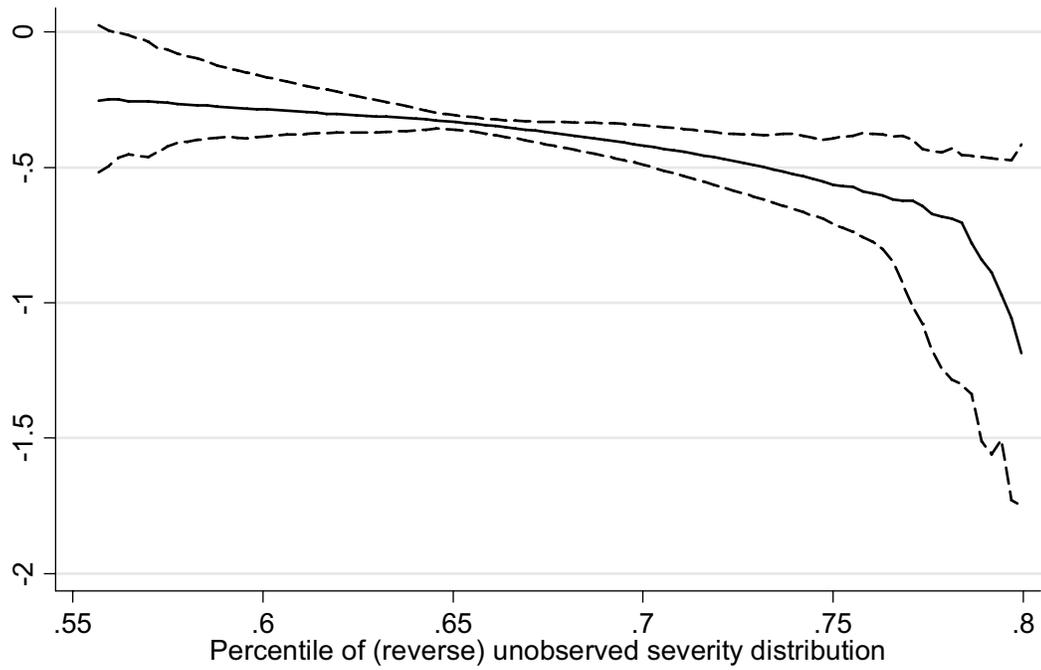
Source: DIODS data for 2005 and 2006.
95% confidence intervals shown with dashed lines.
Bandwidth is 0.084.