University of Michigan Retirement Research Center

Working Paper
WP 2010-241

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

Nicole Maestas, Kathleen Mullen and Alexander Strand

MRRC

Project #: UM10-04

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

Nicole Maestas
RAND

Kathleen Mullen
RAND

Alexander Strand
Social Security Administration

December 2010

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

Acknowledgements

This work was supported by a grant from the Social Security Administration through the Michigan Retirement Research Center (Grant # 10-M-98362-5-01). The findings and conclusions expressed are solely those of the author and do not represent the views of the Social Security Administration, any agency of the Federal government, or the Michigan Retirement Research Center.

Regents of the University of Michigan
Julia Donovan Darrow, Ann Arbor; Laurence B. Deitch, Bingham Farms; Denise Ilitch, Bingham Farms; Olivia P. Maynard, Goodrich; Andrea Fischer Newman, Ann Arbor; Andrew C. Richner, Grosse Pointe Park; S. Martin Taylor, Gross Pointe Farms; Katherine E. White, Ann Arbor; Mary Sue Coleman, ex officio

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

Abstract

We present the first estimates of the causal effects of SSDI receipt on labor supply that are generalizable to the entire population of program entrants in the present day system. We take advantage of a unique workload management database to match Social Security Disability Insurance (SSDI) applicants to disability examiners, and use natural variation in examiners' allowance rates to estimate the labor supply effects of SSDI. Because applicants are randomly assigned to examiners (conditional on observable characteristics), examiner specific allowance rates can be used to instrument for the allowance decision in a labor supply equation contrasting denied vs. allowed applicants. We find that the labor force participation rate of the marginal entrant would be on average 21 percentage points greater in the absence of SSDI benefit receipt. His or her likelihood of engaging in substantial gainful activity as defined by the SSDI program would be on average 13 percentage points higher, and he or she would earn \$1,600 to \$2,600 more per year on average in the absence of SSDI benefit receipt. The marginal entrant is likely to have a mental impairment, be young, and have low pre-onset earnings. Importantly, the disincentive effect varies across individuals with impairments of different degrees of unobservable severity, ranging from a low of 10 percentage points for those with more severe impairments to a high of 60 percentage points for entrants with relatively less severe impairments.

Authors' Acknowledgements

We thank Josh Angrist, David Autor, Raj Chetty, Kirk Doran, Lee Lockwood, Erin Johnson, Dayanand Manoli, Heather Royer, Heidi Williams, and participants in the 2010 MRRC Researcher Workshop, the 2010 All-CA Labor Economics Conference at UC-Santa Barbara, and seminar participants at the Center for Business and Public Policy at the University of Illinois, Urbana-Champaign, for helpful comments and suggestions. This research was supported by a grant from the U.S. Social Security Administration (SSA) through the Michigan Retirement Research Center (MRRC). The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA or any agency of the Federal Government.

1. Introduction

Social Security Disability Insurance (SSDI) is intended to replace lost income for individuals suffering substantial long-term losses in earnings capacity due to the onset of a disability. Specifically, it is targeted at individuals who suffer a long-lasting impairment that *prevents* work. Yet, the dramatic rise in the SSDI caseload over the last two decades and the attendant compositional shift toward disabilities with greater diagnostic uncertainty – along with medical advancement and mandated workplace accommodation for individuals with disabilities – have led many to question whether some SSDI beneficiaries could actually work if they wanted to do so. Recent and proposed policies have attempted to encourage work among beneficiaries. For example, the Ticket to Work Incentive and Work Incentives Improvement Act (Ticket Act) in 1999 provided vouchers (or "tickets") for individuals interested in obtaining vocational rehabilitation and employment services. In addition, the Ticket Act directed the Social Security Administration (SSA) to study the impact of reducing the program's very high implicit tax rate on earnings by gradually phasing out benefits as earnings rise instead of the current policy of suspending the entire benefit when earnings exceed a disregard level.

A key parameter for understanding the likely effects of policies to encourage work among SSDI beneficiaries is the disincentive effect of SSDI on the labor force participation and earnings of recipients. To obtain a causal estimate of the disincentive effect, one needs an estimate of the counterfactual earnings of SSDI beneficiaries. In this paper, we construct a causal estimate of the disincentive effect of SSDI by comparing the subsequent labor supply of otherwise similar applicants who were initially 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. To do this, we use a relatively new administrative data set containing 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. The ultimate allowance decision may coincide with the initial determination or, if the initial determination was appealed, the decision rendered on appeal. We link the applications data to administrative earnings records and examine labor supply outcomes 2 to 3 years from the date of the initial determination. By waiting at least two years to measure labor supply outcomes, we address the possibility that some initially denied applicants may intentionally keep their earnings low during the appeals process.

Our instrumental variable enables us to estimate a local average treatment effect (LATE) for the *marginal entrant,* where "treatment" refers to DI receipt. That is, our estimate is relevant for "compliers" or those whose treatment status is changed because they were assigned to a particular examiner. We find that labor force participation of the marginal entrant would be on average 21 percentage points greater in the absence of SSDI benefit receipt. His or her likelihood of engaging in substantial gainful activity as defined by the SSDI program would be on average 13 percentage points higher, and he or she would earn $1,600 to $2,600 more per year on average in the absence of SSDI benefit receipt. We also investigate the characteristics of the marginal SSDI entrant, on the margin of an initial allowance in the present system; we find that the marginal entrant is more likely to have a mental impairment, is younger, and has pre-disability onset earnings in the lowest earnings quintile.

In a world with homogeneous treatment effects conditional on application, LATE generalizes to the average treatment effect (ATE) – the causal effect of DI receipt on the labor supply of *all* applicants – and the average effect of treatment on the treated (TT) – the causal effect of DI receipt on *beneficiaries* in particular. The unique nature of our instrument, combined with data on the universe of applicants, presents an unprecedented opportunity to test for heterogeneous treatment effects and evaluate the generalizability of previous estimates in the literature. First, we estimate local average treatment effects for subpopulations defined by observable characteristics such as impairment type, age, and prior earnings, and find evidence of a heterogeneous treatment effect. Next, we push this line of investigation further and test for variation in the treatment effect on *unobservable* dimensions, namely unobserved severity of impairment. As we show later, variation in our continuous instrument (examiner allowance propensities) traces out the treatment effect as the allowance threshold is set at different points in the distribution of unobserved severity.

We show that the *marginal treatment effect* (MTE) is increasing in magnitude in the SSDI allowance rate; that is, the disincentive effect of SSDI rises as the allowance threshold is lowered and applicants with less severe impairments are allowed on the program. When the allowance threshold is particularly high, such that the allowance rate is ten points lower than in the present system, the marginal entrant has a high-severity impairment and SSDI causes a modest 10 percentage point reduction in labor force participation. As the allowance threshold is relaxed, the disincentive effect grows. When the allowance threshold is particularly low, such that the allowance rate is 10 points higher than at present, the marginal entrant has an impairment of lesser (unobserved) severity, and SSDI causes a 60 percentage point reduction in labor force participation. Our estimates point to markedly greater residual work capacity among applicants with less severe impairments, and indicate large potential labor supply effects from policies aimed at tightening or relaxing access to SSDI benefits.

Not surprisingly given its obvious importance, the labor supply disincentive effect of SSDI has received a great deal of attention in the economics literature. Bound (1989) first proposed using denied applicants as a comparison group for SSDI beneficiaries. He argued that, in observational data, the labor supply of denied applicants could be used to estimate an upper bound of the effect of SSDI because, although allowed and denied applicants are similar in many respects, on average denied applicants should suffer from less severe impairments than allowed applicants. He estimated that labor force participation of SSDI beneficiaries would be at most 34 percentage points higher in the absence of SSDI benefits.[1] von Wachter, Song and Manchester (2010) found that this upper bound has been relatively stable over time, despite changes in the composition of SSDI applicants. Gruber and Kubik (1997) took this line of approach a step further and exploited state-level variation in a tightening of access to SSDI benefits in the late 1970s to examine the reduced form relationship between denial rates and labor supply. They found that a 10% increase in the state denial rate led to a statistically significant increase in labor force participation of 2.8%.

Two recent studies also estimate an average causal effect of SSDI, but for potentially non-representative subpopulations of SSDI entrants. Chen and van der Klaauw (2008) used exogenous variation in the SSDI denial rate resulting from a little known program rule that effectively relaxes access to SSDI discontinuously at age 55. They found that labor force participation of older beneficiaries would be at most 20 percentage points higher in the absence

[1] Note that this is not the same as saying fewer than 34 percent of beneficiaries would work if the SSDI program did not exist. This is only the case if the act of applying for SSDI benefits has no impact on labor supply or earnings. Since applying for SSDI requires one to reduce one's earnings below a low threshold in order to qualify for benefits, individuals may lose skills while they wait for a decision or have a hard time obtaining a job after a significant absence from the labor market.

of SSDI. Most recently, and concurrent with our study, French and Song (2009) 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 labor force participation rate of applicants allowed at the ALJ hearing level would be at most 14 percentage points higher if they had instead been denied. While the internal validity of the Chen-van der Klaauw and French-Song studies is high, owing to the use of quasi-experimental variation, their generalizability to the full population of SSDI entrants is unknown. Importantly, if the causal effect of SSDI on labor supply is not constant over the population of SSDI entrants, then these two estimates are local average treatment effects that depend critically on the population from which the sample is drawn and the margin affected by the instrumental variable. In contrast, our estimates of the average causal effects of SSDI generalize to the entire population of SSDI entrants.

The paper is structured as follows. Section 2 provides a short overview of the SSDI program with a focus on rules relevant for program entry and labor supply of program participants. Section 3 provides a description of our empirical strategy and Section 4 describes the administrative data sets used to construct our samples. Section 5 presents the results of the first stage of the instrumental variables estimation. Section 6 presents estimates of the local average and marginal treatment effects of SSDI on labor supply and earnings of beneficiaries. Section 7 concludes.

2. Background on SSDI

The SSDI program defines disability as the “inability to engage in any substantial gainful activity (SGA) by reason of any medically determinable physical or mental impairment(s) which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months." 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 $1,000 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).

Individuals may apply for SSDI benefits at their local field office, which screens out those who are currently engaging in SGA. These are known as "technical denials" and they do not receive a medical review. The remaining applications are forwarded to the state Disability Determination Services (DDS) office, where cases are assigned to disability examiners. The 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 (RFC), or ability to work given their physical and/or mental impairments.[2]

The SSDI program employs a five-step review process (see Figure 1) in order to determine whether applicants are unable to perform SGA given the severity of their impairment and/or vocational background (age, education and work experience). Two types of allowances are made: medical allowances (at step 3, for individuals with specific impairments deemed severe enough to warrant allowance into the program regardless of vocational background) and vocational allowances (at step 5, for individuals with severe enough impairments to prevent them from performing SGA given their vocational background). Vocational allowances are made in

[2] According to conversations with a former DDS examiner, medical consultants are used for fewer than half of all applications.

consultation with a medical-vocational grid which provides guidance for allowance decisions based on residual functional capacity, age group, education and work experience (i.e., skilled, unskilled). Practically, the grid generates a large increase in the allowance rate at age 55 (see Chen and van der Klaauw, 2008, for more details).

Denied applicants can appeal their initial determination within 60 days by applying for reconsideration, which sends the case back to the original DDS office. 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. Applicants denied by an ALJ have additional opportunities to appeal to an SSA Appeals Council (AC) and finally to Federal Court. In our sample, roughly one-third of applicants are allowed in the initial determination (divided approximately evenly among the two types of allowances), and just under two-thirds of applicants are ultimately awarded SSDI benefits. An audit study performed by the Office of the Inspector General (2008) found that in 2006 the average processing time for cases in the initial determination phases was 131 days, or just over a 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. Just under a third of cases made it to the ALJ level, where approximately 61% of denials were overturned (GAO 2004). Less than 5% of cases progressed to the AC level and less 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 while they wait for a decision or for a period of at least five months (whichever comes first) before they are entitled to receive benefits. Back payments are made to beneficiaries whose applications take longer than five months to process. 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. Finally, upon reaching the Social Security Full Retirement Age, SSDI benefits are automatically converted to Social Security and the SGA earnings restriction is lifted.

3. Empirical Strategy

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

$$y_i = X_i\beta + \gamma DI_i + u_i, \quad (1)$$

where y_i is earnings (or labor force participation) 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 allowed (i.e., is currently an SSDI beneficiary), and u_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 receipt. 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) earnings loss associated with the individual’s impairment, and which is uncorrelated with any remaining idiosyncratic element ε_i. Then in the regression above $u_i = -s_i + \varepsilon_i$, 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), assuming $\gamma < 0$ and severity positively correlated with SSDI receipt, OLS overestimates the magnitude of the coefficient on *DI* and provides an upper bound on the labor supply effect of SSDI.

In an ideal world, SSDI is 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 severity – which the examiner observes in greater detail than the econometrician, through medical records and test results – and characteristics of the examiner assigned to the case, such as previous experience or personal perceptions/tastes. Let

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

where σ_j denotes a systematic component of examiner judgment that leads some examiners to over- or underestimate severity of applicants on average and ν_{ij} denotes an i.i.d. idiosyncratic element uncorrelated with either severity or σ_j. Then the assignment rule becomes

$$DI_i = 1(s_i + \nu_{ij} > X_i\beta - SGA - \sigma_j), \qquad (2)$$

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

$$P_j(X_i) = 1 - F_{s+v}(X_i\beta - SGA - \sigma_j), \quad (3)$$

where F_{s+v} denotes the cumulative distribution function for unobserved severity plus idiosyncratic examiner error, v_{ij}, which is assumed to be identically and independently distributed within and across examiners. 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. In particular, we instrument for SSDI receipt in Equation 1 using assigned examiner's propensity P_j.

Since appeals are likely to be non-random, we adopt an intention-to-treat framework, in which assignment to disability examiner in the *initial* determination process is used to instrument for the ultimate award decision. Disability examiners are characterized by their caseload initial allowance rate, and initial determinations may differ from final determinations as a result of appeals. Note that since the appeals process is one-sided – that is, only denied applicants appeal their decision and have the possibility of having it overturned –the instrument is likely to have a weaker effect on individuals assigned to "tough" examiners, or those with low σ_j's.

In the case where treatment effects are heterogeneous (i.e., replace γ in Equation 1 with γ_i to allow it to vary across individuals) then an instrumental variables (IV) estimator using initial allowance propensities as an instrument for ultimate allowance decisions will estimate a local average treatment effect (LATE) where the causal effect of DI receipt on labor supply decisions is averaged over "compliers," or appliants whose treatment status is changed as a result of assignment to a particular examiner. These include individuals who are assigned to relatively "lenient" examiners, as well as those who would not appeal a decision if not awarded benefits –

i.e., those whose expectation of a reversal is low (individuals with lower severity) or those for whom an appeal is difficult to initiate (e.g., if it is difficult for them to remain out of the labor force for two or more years). Note that our instrumental variable is not informative for "always takers" – those who would be allowed, whether initially or upon appeal, even if assigned to the stricted examiner – or "never takers" – those who would be denied even if assigned to the most lenient examiner. Thus, we estimate the average causal effect of SSDI for the *marginal entrant* who is just on the margin of being allowed or denied under the current system.

Importantly, if the treatment effect is heterogeneous, then the population on which the estimate is based is key to understanding the extent to which the estimate generalizes to the broader SSDI population. In this paper, we use data on the universe of applicants for whom an initial medical determination is made, whereas other studies use data on specific subsets of the applicant population. For example, Chen and van der Klaauw's (2008) estimate is relevant for individuals on the eve of turning age 55 who are also on the margin of a vocational allowance – that is, individuals in the middle of the severity distribution, once obvious denials (at step 2) and obvious allowances (at step 3) have been removed from the application pool.

By contrast, French and Song (2009) estimate labor supply effects for the set of individuals denied in both the initial determination and reconsideration stages who appeal their decision to an ALJ. As a result, their sample excludes individuals assigned to relatively lenient DDS examiners, who were awarded disability benefits, as well as individuals assigned to relatively strict DDS examiners, who were denied benefits but who did not appeal the initial determination. This last group includes those whose expectation of a reversal is low (individuals with lower severity) or those who have a higher opportunity cost of remaining out of the labor force for two or more years). Practically, this suggests that their estimate should be lower than the average treatment effect for applicants on the margin of allowance at the initial determination stage, as their sample includes individuals with higher severity and lower opportunity costs of appeal (e.g., lower potential earnings).[3] At the same time, their estimate is relevant for policies aimed at reforming the appeals process *holding fixed the current DDS determination process.*

Because our data set includes the entire population of program applicants, we can test for a heterogeneous treatment effect on the basis of both observable and unobservable characteristics. We test for heterogeneity on the basis of observable characteristics by estimating the IV model above separately for groups of applicants defined by impairment type, age, and pre-disability earnings quintile.

We test for heterogeneity on the basis of unobservables by noting that the continuous nature of our instrument enables us to estimate the marginal treatment effect (MTE) γ_i (Heckman, Urzua, Vytlacil, 2006). Note that the MTE is the treatment effect for the marginal entrant with severity $s_i = X_i\beta - SGA - \sigma_j - v_{ij}$. Then the MTE estimated at different predicted probabilities of SSDI receipt generated by the instrument $P_j(X_i)$ implicitly traces out the marginal treatment effect at different points in the distribution of unobserved severity. In particular, we can interpret the MTE computed at low values of predicted SSDI receipt as the MTE for high severity applicants, and vice versa. Practically, computation of the MTE can be accomplished by computing numerical derivatives of a smoothed function relating mean labor supply outcome, y_i, to the predicted probability of SSDI receipt.

Finally, while it is reasonable to expect that the disincentive effect, γ_i, may be related to unobserved severity, s_i, these are two distinctly different concepts. Whereas s_i denotes the

[3] In the meantime, these individuals may also have lost a substantial amount of human capital while remaining out of the labor market, so that denied applicants at this stage may have an even harder time finding a job than the average denied applicant in the aftermath of a negative decision.

expected earnings difference between two otherwise identical individuals, one of whom is disabled while the other is in perfect health, by contrast γ_i denotes the expected earnings difference between two individuals *with the same severity of impairment*, one of whom is receiving disability benefits while the other is not. A priori it is not clear what the expected relationship between γ_i and s_i should be, but if we interpret the disincentive effect of SSDI as a measure of the residual work capacity of applicant *i*, then a reasonable prior is that γ_i is a decreasing function of s_i; in other words, the magnitude of the disincentive effect is increasing in s_i.

4. 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."[4] Because the DIODS is meant to track disability claims for workload management purposes, it contains alphanumeric codes that identify disability examiners uniquely within DDS. This allows us to construct a measure of each examiner's propensity to allow SSDI applicants. 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 primary claimants (i.e., excluding dependents) between the ages of 18 and 64 who were assigned to examiners handling at least 10 cases in the observed time frame.[5] We also exclude cases assigned to examiners with more than

[4] http://www.socialsecurity.gov/disability/data/ssa-sa-mowl.htm
[5] Unfortunately, we cannot distinguish between full- and part-time examiners, nor can we observe examiner tenure.

900 cases in the two year period.[6] We link the applicant data to SSA's Master Beneficiary Record (MBR) in order to identify which applicants ultimately received SSDI benefits (by the end of 2009), which may differ from the initial decision due to the appeals process.[7]

To examine past and subsequent labor supply, we link the applications data to administrative annual earnings records between 1995 and 2008 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 2-3 years after the initial decision.[8] We examine labor supply and earnings at least two years after the initial decision to allow most of the denied applicants to clear the appeals process. Finally, we link the sample to the SSA's Numerical Identification System (Numident) file, which contains dates of death for individuals who died during our sample period. We restrict the analysis samples to individuals who were alive through the end of the calendar year two or three years from the initial decision.

Figure 2 presents the number of applications to SSDI between 1999 and 2008 based on the 2009 Annual Statistical Report for the Social Security Disability Insurance Program. In 2005 and 2006, approximately 2.1 million applicants per year submitted claims for SSDI benefits, and roughly 600,000 of these resulted in technical denials. The remaining 1.5 million applications were sent to a DDS for a medical/vocational determination. After imposing the sample

[6] According to conversations with a DDS examiner, a full time examiner can expect to handle approximately 300 cases per year. We exclude outliers, i.e., those who are observed to handle more than 150% of the average caseload. This restriction removes applications for 25 examiners.

[7] Also, applicants may file additional claims. As a result, and because we use observed SSDI benefit receipt rather than observed allowance upon appeal, our data will have a higher proportion of allowances than some published statistics based on observed appeal outcomes.

[8] We observe earnings up to two years after the initial decision for the full sample, and up to three years later for 2005 applicants.

restrictions discussed above and removing duplicates, we obtain a sample of just under 2.4 million observations for 2005 and 2006 combined.

Table 1 presents summary statistics for the applicant data set, overall and by case disposition. We divide the sample into four groups: 1) initially allowed, 2) initially denied, but allowed on appeal, 3) initially denied, and denied on appeal, and 4) initially denied, and not observed to appeal.[9] Approximately 38% 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 overturned. Fifty-nine percent of applications are for one of two types of impairments: musculoskeletal (37%) or mental disorders (22%). 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.

Note that the four groups are distinctly different from one another in terms of observable characteristics. For example, ultimately denied applicants resemble those who were initially denied but allowed on appeal in terms of body system affected by the impairment (e.g., musculoskeletal, mental), but differ substantially from the initially allowed. Yet, ultimately denied applicants tend to have less severe impairments (as determined in the initial review process); that is, they were generally more likely to be denied in the earlier stages of the review process than those awarded SSDI benefits on appeal. Note that applicants who were initially denied and declined to appeal their decision were most likely to have been rejected at step 2 of the determination process. Moreover, denied applicants tend to be younger and have lower initial earnings than both initially allowed beneficiaries and those allowed on appeal.

[9] 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.

Figure 3 provides a first look at the labor supply and earnings of applicants before and after their initial determination. We use two measures of labor supply: 1) labor force participation – defined as earning more than $1,000 in a given year;[10] and 2) engaging in SGA (where the SGA threshold is that defined by SSA in a given year).[11] We also present mean earnings before and after the initial decision for each group. All earnings amounts are presented in 2008 dollars. From Figure 3A, we see that applicants allowed on appeal are very similar to applicants allowed in the initial determination in terms of their labor force participation before the initial determination, with a labor force participation rate of approximately 90% between 3 and 5 years before the initial decision. Both groups' labor supply drops significantly one year after the initial decision, although initially denied applicants are about five percentage points more likely to work than initially allowed applicants. By two years after the decision, however, the two groups are virtually indistinguishable, with labor force participation rates around 14% between two and three years after the initial decision.

By contrast, ultimately denied applicants have lower labor force participation to begin with – between 73% and 80% three to five years before the initial determination – and significantly higher (albeit reduced) participation rates of around 50% after the initial determination.[12] Although denied applicants who appeal (unsuccessfully) and those who decline to appeal have similar pre-decision labor force participation rates, their post-decision labor force participation differs substantially. Specifically, only about 40% of individuals who appealed unsuccessfully are working three years after their initial determination. Given average

[10] 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 labor force participation to only "meaningful" participation in the labor market.

[11] Note that, since 2000, the SGA threshold has been indexed to average annual wages. However, prior to 2000, the threshold was set nominally and raised infrequently. The SGA threshold was $500 in the 1990s and raised to $700 in 1999.

[12] Recall from Table 1 that approximately 70% of those ultimately denied benefits did not appeal the initial determination.

cumulative processing times from the start of the process, few of these applicants are likely to be still awaiting a decision (and hence suppressing labor supply) particularly when the outcome is measured in the third year after the initial decision. Rather, it likely reflects selection into the appeals process; that is, individuals with less severe impairments are more likely to opt out of the appeals process.

Moreover, note that the difference in prior labor force participation between ultimately allowed and denied applicants is 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 for 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 $11,280 per year in 2008) (see Figure 3B) and average earnings (Figure 3C). Interestingly, Figure 3C suggests there is little difference in prior labor market history between denied applicants who choose to appeal and are unsuccessful, and those who decline to appeal.

Consistent with the earlier evidence from Bound (1989), Figure 3 suggests that labor force participation rates among allowed applicants would be at most 35 percentage points higher in the absence of SSDI, assuming allowed applicants suffer more severe impairments (i.e., with greater losses in earning power) than denied applicants. Similarly, allowed applicants would be at most 21 percentage points more likely to engage in SGA, and would earn at most $8500 per year, if they did not receive SSDI benefits. Regression analysis would allow one to correct these estimates for differences between the two groups in observable factors such as prior earnings, but it would not be able to correct for differences in *unobservable* factors such as severity of the impairment since this is positively correlated with SSDI receipt. To circumvent this problem, we propose using exogenous variation in examiner-specific allowance rates to instrument for the allowance decision in a regression of labor supply (earnings) on SSDI receipt.

5. Initial Outcomes in the Disability Determination Process

Even though maintaining consistency in the disability determination process has been a focus of SSA for a number of years, some variation in screening rates is likely inevitable. The increasing complexity of medical practice and heavy workloads of DDS examiners – combined with the inherent judgment involved when assessing subjective factors such as the degree to which a given impairment inhibits an applicant's ability to work – imply that some variation in allowance rates across examiners is natural. Since consistency in disability awards is important for ensuring accuracy in the dissemination of benefits to the intended recipients, SSA has sponsored a number of studies examining variation in allowance rates across DDS examiners, states and adjudicative levels. In a seminal study, Nagi (1969) commissioned an expert panel to perform external audits on a sample of SSDI applications and found that the panel agreed with the original award decision in just under 70% of cases. In a similar study, Gallicchio and Bye (1981) replicated a sample of claims decisions both within and between states, and found that within-state disagreements were in many cases as large or larger than disagreements between states.

Our approach exploits *systematic* variation in allowance rates across individual examiners within the same state DDS office. To examine whether there is sufficient variation across examiners, we construct a measure of initial allowance rates among cases decided by a given DDS examiner between 2005 and 2006. Note that in our sample, 90% of examiners are observed to handle more than 29 applications, and the median observed caseload is 215, so there should be little concern that our signal-to-noise ratio is low.[13] Figure 4 presents a histogram of examiners' deviation from the overall allowance rate (within DDS office), both unadjusted and regression-adjusted for differences in case mix. Case controls include fraction of cases in a given age group, diagnosis code or month, as well as a variable measuring average prior earnings of applicants assigned to a given examiner. As expected, adjusting for case mix reduces the standard deviation, from 0.0968 to 0.0668, and removes the thick right tail. Controlling for a rich set of covariates, one-third of examiners have allowance rates more than 7 percentage points above or below the average allowance rate for a given case type, and 5% of examiners have allowance rates more than 13 percentage points above or below the average.

Figure 5 is a graphical depiction of our instrumental variables estimation strategy. We plot smoothed SSDI receipt and labor force participation rates, respectively, of applicants by examiner's residualized (regression-adjusted) initial allowance rate, with 95% confidence interval bands, estimated via local quadratic regression. SSDI receipt is increasing in the residualized initial allowance rate, but rises more slowly than the allowance rate on account of the appeals process. An increase in the examiner allowance rate of 10 percentage points above the mean, yields an increase in SSDI receipt of less than half that amount because the system now initially allows many individuals who, had they been initially denied, would have obtained benefits on appeal. At the same time, the labor force participation rate of applicants two years after the initial decision falls as the examiner's allowance propensity rises, from 28% among individuals assigned to tough examiners to 24% among individuals assigned to lenient examiners. Moreover, the rate at which labor force participation falls is related to the rate at

[13] In simulations (available on request), we found that observing 10 applications per examiner was sufficient to detect examiner's "true" allowance propensities +/- 0.2 with 95% confidence. Furthermore, observing more than 30 applications per examiner yielded no real increases in precision (+/- 0.1 with 95% confidence).

which SSDI receipt increases; where the relationship between SSDI receipt and the initial allowance rate is flat, labor force participation is also flat and only begins to decrease as DI receipt increases.

For the regression analysis, in order to avoid biasing measurement of examiner propensities using the applicant's own decision, we construct the following individual-specific instrument for DI receipt:

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

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 further below, EXALLOW should only be positively correlated with the applicant's own award decision if there exists an underlying examiner-specific threshold for allowance decisions. Table 1 presents summary statistics for EXALLOW by disposition of the case. Note that individuals who were initially allowed were assigned to examiners with average initial allowance rates of 40% *excluding themselves*, while individuals who were initially denied were assigned to examiners with average initial allowance rates of 35%, regardless of the ultimate disposition of the case.

In order for EXALLOW to be a valid instrument for DI receipt, applicants' assignment to DDS examiner must be uncorrelated with unobserved severity conditional on observed characteristics. This amounts to an assumption of *conditional* random assignment to DDS examiner within DDS. That is, examiners may specialize in cases within a particular type of impairment (e.g., mental disorders) or age group (e.g., older workers), but within these cases examiners do not divide their caseloads by low or high severity. This is plausible for several reasons. First, little is known about cases when they are assigned and it is unlikely that a first glance would reveal the severity of the impairment beyond general type, which is observable. Second, even if cases are non-randomly assigned within impairment type it is likely that some (e.g., more experienced) examiners might specialize in "difficult" cases, but these cases are those on the *margin* of allowance, whereas "easy" cases come from either extreme of the distribution (very low or high severity). Finally, anecdotal evidence based on conversations with a former DDS examiner suggests that in most offices cases are randomly assigned to examiners.[14]

In our analysis, we control for a rich set of covariates, namely: characteristics of the impairment (body system/diagnosis codes), age, prior labor market history, geography (three-digit zip codes) and seasonality (month of application). Table 2 presents the estimated coefficient on EXALLOW in a regression of individual allowance decisions, with and without covariates. We present results separately for 2005 and 2006, as well as pooled together. All models include DDS dummies to account for stratification of examiners across DDS offices. We present t-statistics in parentheses, where robust standard errors are computed and clustered by DDS examiner. Column 1 presents the coefficient on EXALLOW with no additional covariates. In both years, a 10 percentage point increase in *initial* examiner allowance rate results in just over a three percentage point increase in one's own probability of ultimately receiving DI. Adding a finer measure of geography by grouping individuals by the first three digits of their zipcode (Column 2) does little to change the estimated coefficient, suggesting that examiners do not tend to specialize in a given geographic area within a DDS region. Because including the roughly 900

[14] Note that if we do not completely control for nonrandom assignment based on applicant characteristics that are correlated with labor supply outcomes, this will tend to bias our estimates in the direction of OLS. In that case, our estimates can be interpreted as a tighter upper bound on the labor supply effects of SSDI.

three-digit zip code dummies increased computing time substantially, we dropped them for the remaining analysis.[15]

The next two columns (3-4) add body system codes and diagnosis codes, respectively. Both of these sets of variables have an impact on the coefficient on EXALLOW, suggesting that some type of specialization according to impairment type is practiced across at least some DDS offices. This may be deliberate (e.g., an examiner tends to focus on cases with mental disorders) or indirect (e.g., if cases involving mental disorders are more difficult to process on average, and as a result they are assigned to more experienced examiners). In any case, this suggests that conditioning on impairment type is critical to our identification strategy. Finally, controlling for age, average previous earnings and seasonality has very little impact on the estimated effect of the instrument on individual allowance decisions. This suggests that, once one conditions on impairment type, applicants are likely approximately randomly assigned to examiners.

Conditional on observables, examiner allowance rates are strongly correlated with individual allowance decisions: a 10 percentage point increase in examiner allowance rate is associated with a statistically significant ($p<0.0001$) 2.3 percentage point increase in individuals' allowance decisions.[16] This suggests that examiner allowance rates can be used effectively as an instrumental variable for individual allowance decisions.

6. Results

6.1. Effect of SSDI on Labor Supply and Earnings of the Marginal Entrant

[15] Since it is plausible that three-digit zipcode could matter for labor supply outcomes (e.g., if local employment opportunities were correlated sufficiently with SSDI receipt, regardless of their impact on EXALLOW), we ran a subset of models for the labor supply outcomes with three-digit zipcodes included and found that they did not affect those estimates either.

[16] Recall that in a model with one endogenous regressor the t-statistic squared is equal to the incremental F statistic.

Table 3 presents results for the labor supply regressions estimated, respectively, by Ordinary Least Squares (OLS) and Instrumental Variables (IV) using Two-Stage Least Squares where the first stage regression results are presented in the final column of Table 2. We model three outcomes: labor force participation (defined as earning more than $1,000/year), engagement in Substantial Gainful Activity (or SGA; i.e., earning more than $11,280 per year in 2008), and annual earnings. We measure all labor supply outcomes two years after the initial decision, for applicants whose initial determination took place in 2005 or 2006, respectively, and we examine outcomes three years later for those receiving an initial determination in 2005. In most cases, adjusting for differences in observable characteristics between denied and allowed applicants via OLS does little to change estimates of the labor supply effect of SSDI based on comparisons of raw means (not shown). However, adjusting for *unobservable* differences (e.g., in severity of the impairment(s)) has a large impact on the estimated labor supply effect. Consistent with the idea that severity is positively correlated with DI receipt and the idea that OLS estimates an upper bound on the labor supply effect, the OLS estimates are all statistically significantly and substantively larger in magnitude than the IV estimates. The estimates for all three outcomes are relatively stable across decision years, and time after the initial determination.

Labor Force Participation. Table 3 shows that despite their impairments, between 13% and 15% of SSDI beneficiaries participate in the labor market two to three years after entering the program – earning more than $1,000 in a given year. Adjusting for differences in observable characteristics such as impairment type, age or prior labor market history, about half of denied applicants also participate in the labor market – a difference of approximately 34 percentage points (the OLS coefficient). This is almost identical to Bound’s (1989) estimates using data from the 1972 Survey of Disabled and Non-Disabled Adults and the 1978 Survey of Disability of Work. In other words, about half of SSDI beneficiaries would be likely to work if they did not participate in the SSDI program. Moreover, applying our IV identification strategy – which allows us to adjust for differences in *unobservable* characteristics in addition to differences in observable characteristics across the two groups – yields smaller coefficient estimates, between 21 and 22 percentage points after two years and 17 percentage points after three years. We find that once we adjust for differences in unobservable characteristics, about 36.5% of beneficiaries would be likely to work after the onset of disability in the absence of receiving SSDI benefits. Recalling that more than 85% of these same beneficiaries were labor market participants at least three years before their initial determination, this translates to a more than 50% drop in labor force participation due to the onset of disability alone.

Engaging in Substantial Gainful Activity. As noted earlier, the SSA suspends SSDI benefits for individuals earning more than the SGA threshold, which in 2008 was set at $940 per month, or $11,280 per year.[17] However, SSA first gives beneficiaries the opportunity to "test" their ability to return to work by allowing them to engage in SGA for at least twelve months (during the nine-month TWP and three-month Grace Period). Because the TWP and Grace Period are not necessarily consecutive, it is possible to observe SSDI beneficiaries engaging in SGA more than 2 years after their initial determination, and indeed Table 3 shows that around 5% of beneficiaries earn more than the annualized SGA threshold in a given year. [18] Using denied applicants as a control group and adjusting for differences in observable characteristics, the OLS coefficients imply that about 29% (=[0.052+0.235]x100 for 2 years after 2005

[17] Although the SGA threshold is expressed in terms of earnings per month, practically SSA audits annual earnings to flag cases requiring a more intensive audit of monthly earnings.
[18] Note that, because of the earnings restriction, these are unlikely to be the same beneficiaries engaging in SGA in more than one year.

decisions) of SSDI beneficiaries would engage in what SSA considers to be substantial gainful activity two years later in the absence of SSDI benefits. Adjusting for differences in unobservable characteristics using our IV strategy, we find that 17% of SSDI beneficiaries would engage in SGA if their applications had been denied. This represents a more than 30% post-onset drop in the likelihood of having substantial earnings in the absence of SSDI.

Annual Earnings. Finally, we examine average earnings losses associated with SSDI receipt. Table 3 shows that mean annual earnings of SSDI beneficiaries are between approximately $1,800 and $2,000 per year. OLS estimates of the effect of SSDI on average earnings range from approximately $6,500 to $7,300, implying that the average earnings of beneficiaries in the absence of SSDI would be no higher than $9,300 per year – below the 2008 annualized SGA threshold of $11,280. Moreover, IV estimates (between $1600 and $2600) imply that average earnings would be no higher than $4,600 per year, representing a more than 80% post-onset drop in earnings and well below the threshold for substantial gainful activity.

6.2. Treatment Effect Heterogeneity on the Basis of Observable Characteristics

One advantage of our empirical strategy, combined with administrative data, is that it allows us to estimate labor supply effects of DI receipt conditional on observable characteristics such as impairment type, age or prior earnings. Tables 4 and 5 present first stage and second stage results, respectively, for these disaggregated groups. As shown in Table 4, although the instrumental variable is statistically significant in all cases, the strength of the first stage varies substantially across groups, particularly by impairment type. For example, the coefficient on EXALLOW for musculoskeletal cases is 0.162, compared with 0.350 for mental disorders. Doyle (2008) showed that one could use the first stage coefficient estimates in order to characterize the types of individuals most likely to be induced into treatment as a result of the instrumental variable (i.e., the marginal entrant).

Specifically, the relative likelihood that an individual is induced to receive disability benefits as a result of DDS examiner assignment is given by the ratio of the first stage coefficient conditional on a given characteristic relative to the overall first stage coefficient. Thus, a policy change that uniformly increases the probability of allowance at the initial determination stage is 50% more likely to induce an individual suffering from a mental disorder, and 30% less likely to induce an individual with a musculoskeletal impairments. This is not surprising, given that 64% of allowances for mental disorders result from initial determinations, compared with 38% of allowances for musculoskeletal impairments.[19] Similarly, a policy that uniformly increases the probability of allowance at the initial determination stage would be more likely to benefit younger applicants and applicants with low prior earnings. Such compositional changes could have important implications for government spending even if the disincentive effect of disability insurance receipt on labor supply is constant across impairment types. 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 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).

Table 5 presents the estimated effect of disability insurance on labor force participation two years after the initial decision disaggregated across groups. In order to improve the precision of our estimates, we pooled the 2005 and 2006 samples; however, most of our estimates are still

[19] Interestingly, the bulk of musculoskeletal cases are allowed at the reconsideration or appellate stages.

too imprecise to make many definitive conclusions. Only five of the fifteen body system codes yield IV estimates statistically distinct from OLS, and only two of these that are distinct from OLS are statistically distinguishable from zero at the 95% level; not surprisingly, these are musculoskeletal and mental impairments, which make up more than half of the sample. The estimated effect for special senses and speech is -8 percentage points (significant at the 10% level), and respiratory and genitourinary impairments actually yield positive, although not statistically significant, estimates. The remaining impairments are not statistically distinguishable from OLS. There does not appear to be a pattern between the relative magnitudes of the OLS estimates vs. IV estimates for body system codes.

In contrast, dividing the sample into eight age groups produces interesting results. Figure 6 presents the estimates graphically, with 95% confidence intervals for the IV estimates. (Confidence intervals for the OLS estimates are too narrow to be detected visually on the same scale as the IV estimates.) Note that the OLS estimate of the age profile is relatively flat until age 50 and then the magnitude of the labor supply effect declines steeply. By contrast, the IV estimate of the age profile gradually increases in magnitude until around age 45 before beginning its decline, perhaps reflecting fewer employment opportunities for inexperienced or older disabled workers. As a result, the gap between OLS and IV decreases with age, until roughly age 45. Recall the gap between OLS and IV in our model is $[E[s_i \mid DI_i = 1] - E[s_i \mid DI_i = 0]]$, or the difference in average severity of allowed vs. denied applicants. Thus, our estimates imply that, in terms of unobserved severity, denied applicants more closely resemble allowed applicants as they age. Finally, we estimate that disincentive effects of disability insurance are highest for individuals at the very top and very bottom of the (pre-onset) income distribution. For poorer individuals, this may reflect lower opportunity costs of applying for disability insurance. For more skilled (wealthier) individuals, the larger disincentive effect could reflect better (or more accommodating) employment opportunities.

6.3. Marginal Treatment Effects

As noted above, because we instrument for disability insurance receipt using a continuous measure of DDS examiners' allowance propensities, we can trace out the disincentive effect of disability insurance as a *function of unobserved severity* by estimating marginal treatment effects. This exercise is particularly interesting in that it tests for heterogeneous treatment effects on the basis of *unobservables*. If the average treatment effect varies over the distribution of unobserved severity, then estimates of the disincentive effect based on subpopulations of SSDI entrants with different average severity levels are more appropriately regarded as *local* average treatment effects.

Figure 7 shows the MTE over the support of predicted SSDI receipt in the range of 55 percent to 75 percent (with 95% confidence intervals), which encompasses the allowance rate of 65 percent in the present system. At more extreme values of predicted SSDI receipt there are relatively smaller numbers of observations and the instrument lacks precision; when this occurs, the MTE is not identified. The MTE declines monotonically as the SSDI allowance rate increases exogenously; 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. When the allowance threshold is particularly high, such that the allowance rate is ten points lower than in the present system, the marginal entrant has a high-severity impairment and SSDI causes a modest 10 percentage point reduction in labor force participation. As the allowance threshold is relaxed, the disincentive effect grows. At the mean allowance rate for the present system, SSDI causes a 29 percentage point reduction in labor force participation. When the allowance threshold is particularly low, such that the allowance rate is 10 points higher than at present, the marginal entrant has an impairment of lesser (unobserved) severity, and SSDI causes a 60 percentage point reduction in labor force participation. The MTE at the allowance rate corresponding to the present system is somewhat larger than the IV estimate of the local average treatment effect (LATE) of 21 percentage points. This occurs because the IV estimate is in essence a weighted average of the MTEs and depends on the distribution of unobserved severity among applicants, which may be skewed toward higher severity cases.

Our finding that the MTE is declining in unobserved severity (corresponding to a disincentive effect that rises in severity) explains why the French-Song estimate of 14 percentage points is significantly lower than the Chen-van der Klaauw estimate of 20 percentage points and our LATE estimate of 21 percentage points. The French-Song sample of applicants who appeal their initial determination (many of whom are our "always takers," or those individuals who would be allowed regardless of the strictness of their initial examiner) consists of higher-severity individuals; the disincentive effect is smaller for these individuals on account of their lower residual work capacity. While our LATE estimate does not encompass many of those in the French-Song sample, the declining pattern in the MTE suggests that their estimate should be lower than ours. On the other hand, the Chen-van der Klaauw estimate is based on 55 year olds drawn from the middle of the severity distribution and happens to coincide fairly closely with our LATE estimate.

From the perspective of policy, our estimates point to markedly greater residual work capacity among beneficiaries with (unobservably) less severe impairments. As noted earlier, in the present system, the marginal applicant has a mental impairment, is young, and has low pre-disability earnings. A policy with the effect of relaxing access to SSDI benefits in the initial determination phase would draw individuals from this margin, and as the margin moved down the severity distribution, the disincentive effect of SSDI would increase—slowly at first, but increasingly faster.

7. Conclusion

This paper presents causal estimates of the disincentive effect of the SSDI program on the labor supply of program entrants. Our estimates are the first set of estimates of the disincentive effect that are generalizable to the entire population of SSDI entrants 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 labor force participation of the marginal entrant would be on average 21 percentage points greater in the absence of SSDI benefit receipt. His or her likelihood of engaging in substantial gainful activity as defined by the SSDI program would be on average 13 percentage points higher, and he or she would earn $1,600 to $2,600 more per year on average in the absence of SSDI benefit receipt. In the present system, 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 10 percentage points above and below the current system allowance rate, the marginal treatment effect varies from a modest 10 percentage point reduction in labor force participation for individuals with impairments characterized by high unobservable severity, to a 60 percentage point reduction in labor force participation for those with less unobservably severe impairments. 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.

References

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.

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

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

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

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

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

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.

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

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

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), pp. 3-21.

von Wachter, Till, Jae Song and Joyce Manchester (2010). "Trends in Employment and Earnings of Allowed and Rejected Applicants to the Social Security Disability Insurance Program." Mimeo.

Table 1. Summary Statistics

Variable	All Applicants	Initially Allowed	Initially Denied: Allowed on Appeal	Initially Denied: Denied on Appeal	Initially Denied: No Appeal
No. observations	2,380,255	797,447	738,163	249,396	595,249
Percent of sample	100.00%	33.50%	31.01%	10.48%	25.01%
Percent of initial denials	--	--	46.64%	15.76%	37.61%
Initial decision (final step)					
2. Denied - nonsevere impairment	17.16%	0.00%	18.02%	26.44%	35.20%
3. Allowed - met listings	14.66%	43.77%	0.00%	0.00%	0.00%
4. Denied - able to work past jobs	23.64%	0.00%	38.25%	34.31%	32.72%
5. Denied - able to do any work	25.69%	0.00%	43.72%	39.25%	32.06%
5. Allowed - unable to do any work	18.84%	56.23%	0.00%	0.00%	0.00%
Body system code					
Musculoskeletal system	37.02%	25.96%	45.01%	43.19%	39.33%
Special senses and speech	2.46%	3.45%	1.55%	1.85%	2.52%
Respiratory system	3.95%	5.15%	3.37%	3.55%	3.22%
Cardiovascular system	8.20%	8.88%	8.78%	6.93%	7.11%
Digestive system	2.36%	1.87%	2.54%	2.50%	2.72%
Genitourinary impairments	1.46%	3.11%	0.62%	0.45%	0.73%
Hematological disorders	0.35%	0.37%	0.31%	0.30%	0.39%
Skin disorders	0.33%	0.28%	0.31%	0.33%	0.43%
Endocrine system	4.15%	2.51%	5.29%	4.89%	4.62%
Neurological	8.16%	11.16%	7.13%	6.22%	6.21%
Mental disorders	21.59%	27.37%	16.87%	20.21%	20.27%
Malignant neoplastic diseases	3.37%	5.85%	2.10%	1.20%	2.54%
Immune system disorders	2.68%	3.01%	2.71%	2.51%	2.26%
Special/other	3.93%	1.02%	3.41%	5.87%	7.65%
Age at initial decision					
Mean	47.13	50.77	47.38	43.64	43.41
Std. deviation	10.56	10.83	9.23	10.11	12.06
Avg. earnings, 3-5 years before decision					
Mean	22,099.81	27,917.58	22,895.47	15,514.16	16,078.38
Std. deviation	26,878.91	33,116.50	24,122.32	20,853.19	24,465.54
EXALLOW (= initial examiner allowance rate, less own decision)					
Mean	0.37	0.40	0.35	0.35	0.36
Std. deviation	0.10	0.12	0.10	0.10	0.10

Source: DIODS, initial applications decided in 2005-2006

Table 2. First Stage Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2005							
EXALLOW	0.319***	0.305***	0.275***	0.218***	0.212***	0.215***	0.215***
t-stat	37.7	39.66	32.11	28.06	29.6	27.99	28.02
R-squared	0.013	0.018	0.038	0.094	0.155	0.162	0.162
2006							
EXALLOW	0.351***	0.349***	0.308***	0.248***	0.244***	0.246***	0.247***
t-stat	40.82	41.02	36.58	32.34	33.43	32.45	32.67
R-squared	0.014	0.019	0.041	0.098	0.158	0.166	0.166
2005 and 2006							
EXALLOW	0.338***	0.330***	0.294***	0.235***	0.230***	0.233***	0.234***
t-stat	50.14	53.13	42.74	37.97	40.31	37.82	38.42
R-squared	0.013	0.018	0.039	0.096	0.156	0.163	0.164
Control variables included							
3-digit zipcode		X					
Body system codes			X				
Diagnosis codes				X	X	X	X
Age group dummies					X	X	X
Avg. previous earnings						X	X
Month dummies							X
No. control variables	110	1014	125	420	428	429	440

Notes: 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 1,208,898 for 2005 and 1,212,842 for 2006.

Table 3. Effect of SSDI on Labor Force Participation & Earnings

	2005				2006	
	2 Years After Decision		3 Years After Decision		2 Years After Decision	
Outcome	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**
1) Earn >=$1,000/year						
Mean dependent variable | allowed	0.154		0.136		0.139	
Coeff. on ALLOW	-0.338***	-0.215***	-0.349***	-0.170***	-0.322***	-0.206***
	(-326.10)	(-6.75)	(-338.53)	(-5.42)	(-304.66)	(-8.58)
R-squared	0.197	0.183	0.211	0.180	0.188	0.174
2) Earn >= SGA						
Mean dependent variable | allowed	0.052		0.046		0.045	
Coeff. on ALLOW	-0.235***	-0.129***	-0.244***	-0.109****	-0.215***	-0.108***
	(-261.97)	(-6.11)	(-267.38)	(-5.33)	(-227.61)	(-6.09)
R-squared	0.143	0.125	0.157	0.126	0.133	0.112
3) Earnings						
Mean dependent variable | allowed	2,012		1,834		1,775	
Coeff. on ALLOW	-7,150***	-1,634*	-7,343***	-1,562**	-6,529***	-2,564***
	(-140.81)	(-1.71)	(-193.30)	(-2.01)	(-92.92)	(-3.92)
R-squared	0.131	0.094	0.139	0.093	0.134	0.112
No. observations	1,174,852		1,152,585		1,205,403	

Notes: 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, diagnosis codes, age group dummies, avg. previous earnings, and month dummies.

Table 4. Heterogeneity: First Stage Regressions, 2005 and 2006 combined

	No. obs.	Initial Allowance Rate	Ultimate Allowance Rate	Coeff. on EXALLOW	t-Stat.	Relative Likelihood
Body system code						
Musculoskeletal system	881,069	23.5%	61.2%	0.162***	17.06	0.69
Special senses and speech	58,603	47.0%	66.5%	0.439***	20.47	1.88
Respiratory system	93,941	43.8%	70.2%	0.142***	7.33	0.61
Cardiovascular system	195,183	36.3%	69.5%	0.180***	11.64	0.77
Digestive system	56,119	26.6%	60.0%	0.198***	7.13	0.85
Genitourinary impairments	34,835	71.1%	84.3%	0.094***	5.90	0.40
Hematological disorders	8,266	35.2%	63.1%	0.411***	6.79	1.76
Skin disorders	7,933	28.6%	57.3%	0.368***	4.57	1.57
Endocrine system	98,801	20.3%	59.8%	0.071***	2.98	0.30
Neurological	194,148	45.8%	73.0%	0.220***	17.21	0.94
Mental disorders	513,884	42.5%	66.7%	0.350***	37.12	1.49
Malignant neoplastic diseases	80,251	58.1%	77.4%	0.253***	15.76	1.08
Immune system disorders	63,765	37.7%	69.1%	0.247***	10.47	1.06
Special/other	93,457	8.7%	35.6%	0.057**	2.54	0.24
Age at decision						
18-24	78,946	25.6%	40.2%	0.328***	13.61	1.40
25-29	136,461	23.0%	42.6%	0.365***	21.03	1.56
30-34	156,838	22.8%	48.2%	0.332***	18.98	1.42
35-39	211,452	22.4%	52.9%	0.278***	17.61	1.19
40-44	295,526	21.5%	56.6%	0.222***	16.63	0.95
45-49	370,632	22.4%	62.7%	0.195***	16.08	0.83
50-54	399,274	32.0%	72.9%	0.181***	18.56	0.77
55-59	413,497	50.1%	81.6%	0.148***	17.59	0.63
60-64	317,629	57.1%	72.5%	0.279***	24.74	1.19
Avg. prior earnings						
Bottom quintile	476,051	22.1%	44.2%	0.409***	30.62	1.75
2nd quintile	476,051	27.2%	59.0%	0.254***	25.08	1.09
3rd quintile	476,051	31.8%	66.9%	0.199***	20.85	0.85
4th quintile	476,051	38.4%	73.3%	0.181***	19.76	0.77
Top quintile	476,051	48.0%	79.3%	0.158***	19.84	0.68

Notes: t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.10. Control variables include: DDS dummies, diagnosis codes, age group dummies, avg. previous earnings, and month dummies.

Table 5. Heterogeneity: Effect of SSDI on Labor Force Participation (2 Years Later), 2005 and 2006 Combined

			OLS		IV	
	No. obs.	Mean LFP | Allowed	Coeff. On ALLOW	t-Stat.	Coeff. On ALLOW	t-Stat.
Body system code						
Musculoskeletal system	881,069	0.125	-0.349***	-303.13	-0.195***	-4.33
Special senses and speech	58,603	0.209	-0.295***	-69.27	-0.084*	-1.71
Respiratory system	93,941	0.099	-0.294***	-81.97	0.108	0.80
Cardiovascular system	195,183	0.116	-0.316***	-131.96	-0.332***	-4.43
Digestive system	56,119	0.157	-0.344***	-80.89	-0.310**	-2.46
Genitourinary impairments	34,835	0.192	-0.310***	-36.77	0.470	1.38
Hematological disorders	8,266	0.242	-0.322***	-28.43	-0.044	-0.27
Skin disorders	7,933	0.158	-0.374***	-34.70	-0.604***	-3.25
Endocrine system	98,801	0.127	-0.308***	-101.82	-0.396	-1.39
Neurological	194,148	0.124	-0.368***	-140.74	-0.263***	-4.60
Mental disorders	513,884	0.186	-0.317***	-204.78	-0.212***	-7.78
Malignant neoplastic diseases	80,251	0.211	-0.377***	-93.44	-0.316***	-4.67
Immune system disorders	63,765	0.174	-0.314***	-64.24	-0.163	-1.64
Special/other	93,457	0.133	-0.253***	-70.09	-0.636	-1.60
Age at decision						
18-24	78,946	0.363	-0.340***	-88.45	-0.228***	-3.07
25-29	136,461	0.273	-0.357***	-109.80	-0.110**	-2.28
30-34	156,838	0.232	-0.376***	-124.77	-0.221***	-4.28
35-39	211,452	0.205	-0.384***	-171.92	-0.248***	-4.53
40-44	295,526	0.174	-0.370***	-197.96	-0.287***	-5.20
45-49	370,632	0.148	-0.359***	-200.68	-0.314***	-5.79
50-54	399,274	0.127	-0.360***	-204.98	-0.299***	-5.73
55-59	413,497	0.102	-0.323***	-152.58	-0.186***	-3.36
60-64	317,629	0.096	-0.134***	-79.26	-0.107***	-3.33
Avg. prior earnings						
Bottom quintile	476,051	0.148	-0.181***	-141.19	-0.176***	-8.66
2nd quintile	476,051	0.164	-0.337***	-232.30	0.022	0.46
3rd quintile	476,051	0.149	-0.405***	-266.42	-0.062	-1.17
4th quintile	476,051	0.135	-0.444***	-277.59	-0.225***	-4.70
Top quintile	476,051	0.141	-0.429***	-247.30	-0.156***	-2.91

Notes: t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.10. Control variables include: diagnosis codes, age group dummies, avg. previous earnings, DDS dummies and month dummies.

Figure 1. Five-Step Review Process

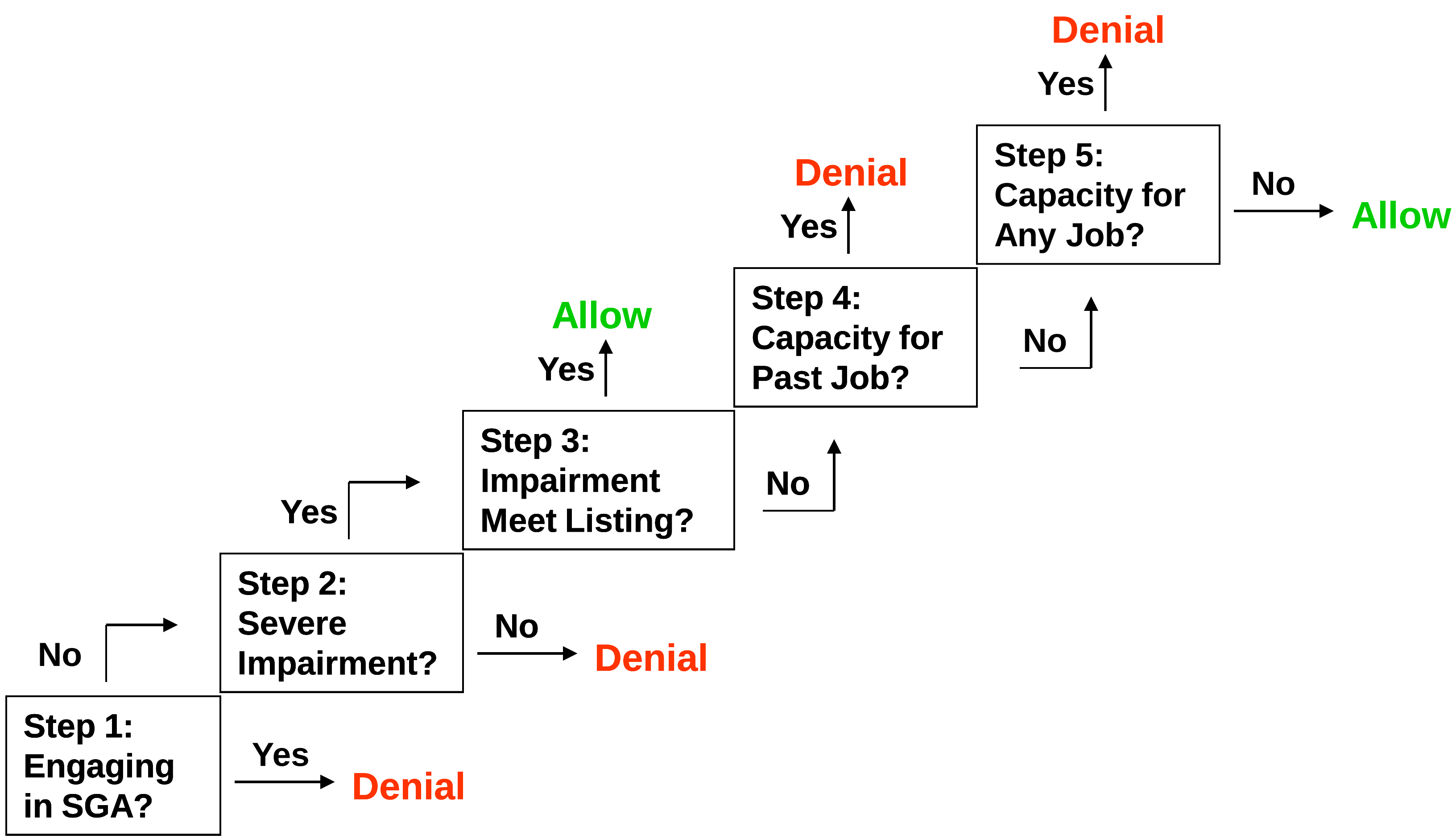
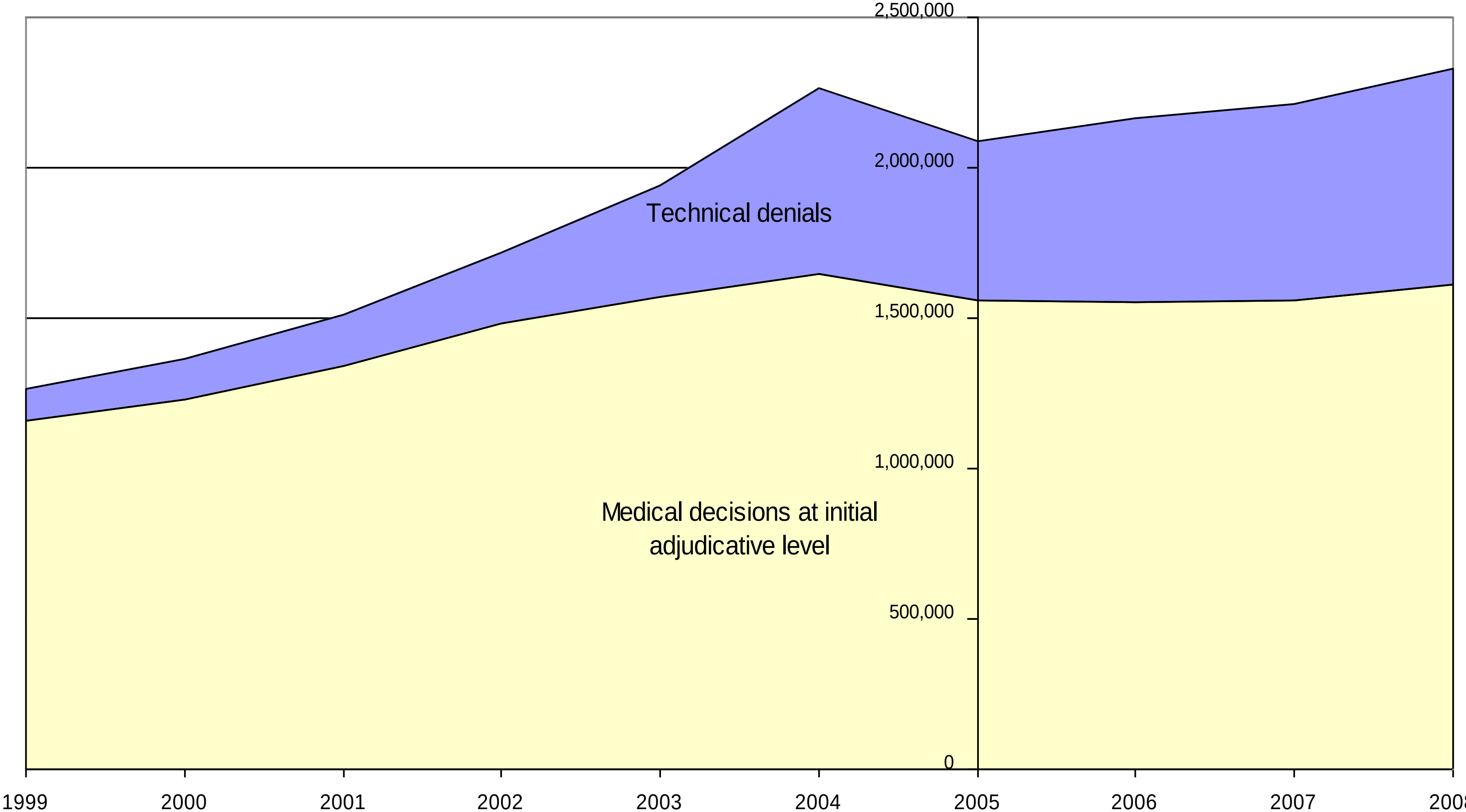


Figure 2. SSDI Applications, 1999-2008



Source: Annual Statistical Report on the Social Security Disability Insurance Program, 2009

Figure 3A. Labor Force Participation Before and After Initial Decision

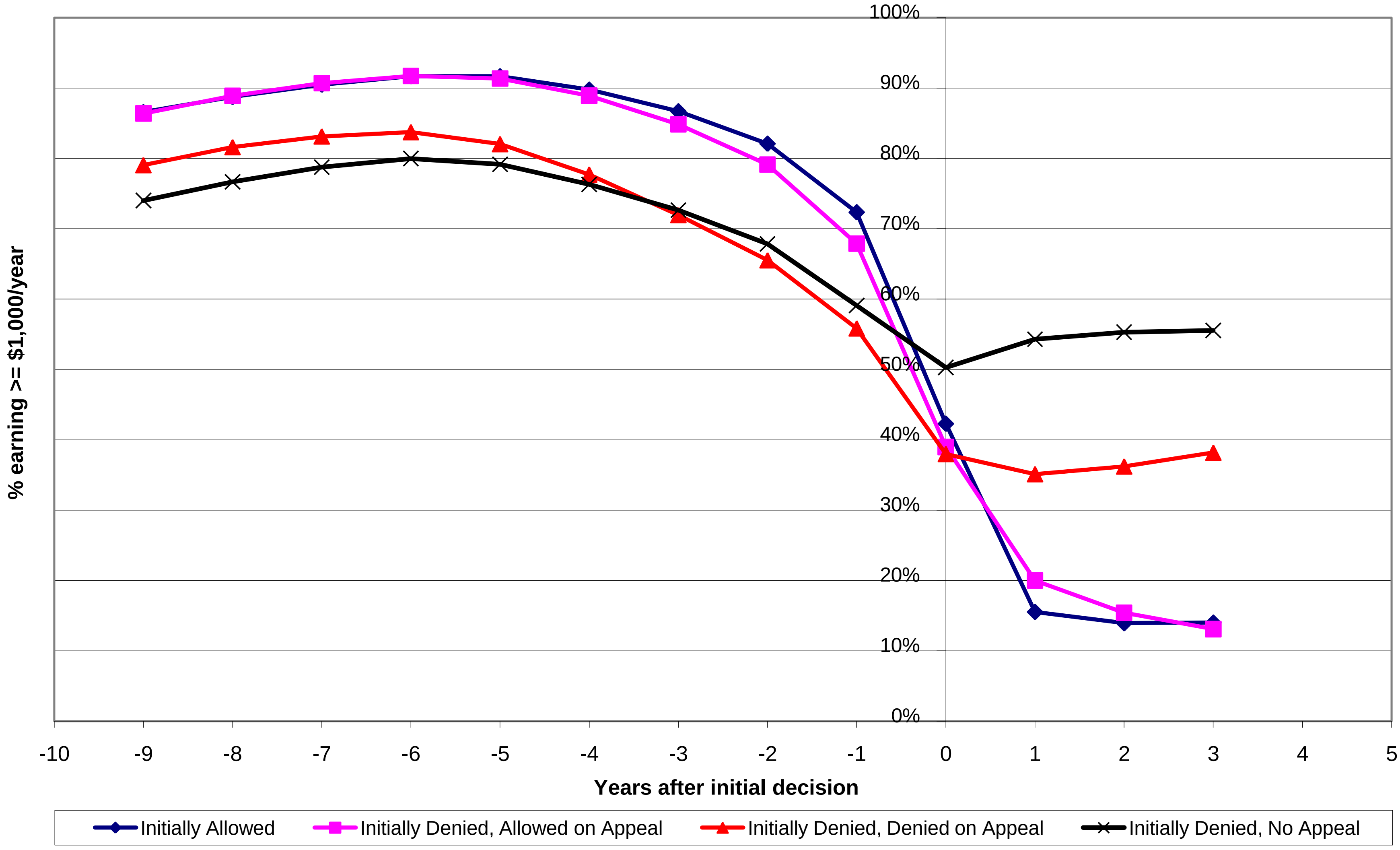


Figure 3B. % Performing Substantial Gainful Activity (SGA) Before and After Initial Decision

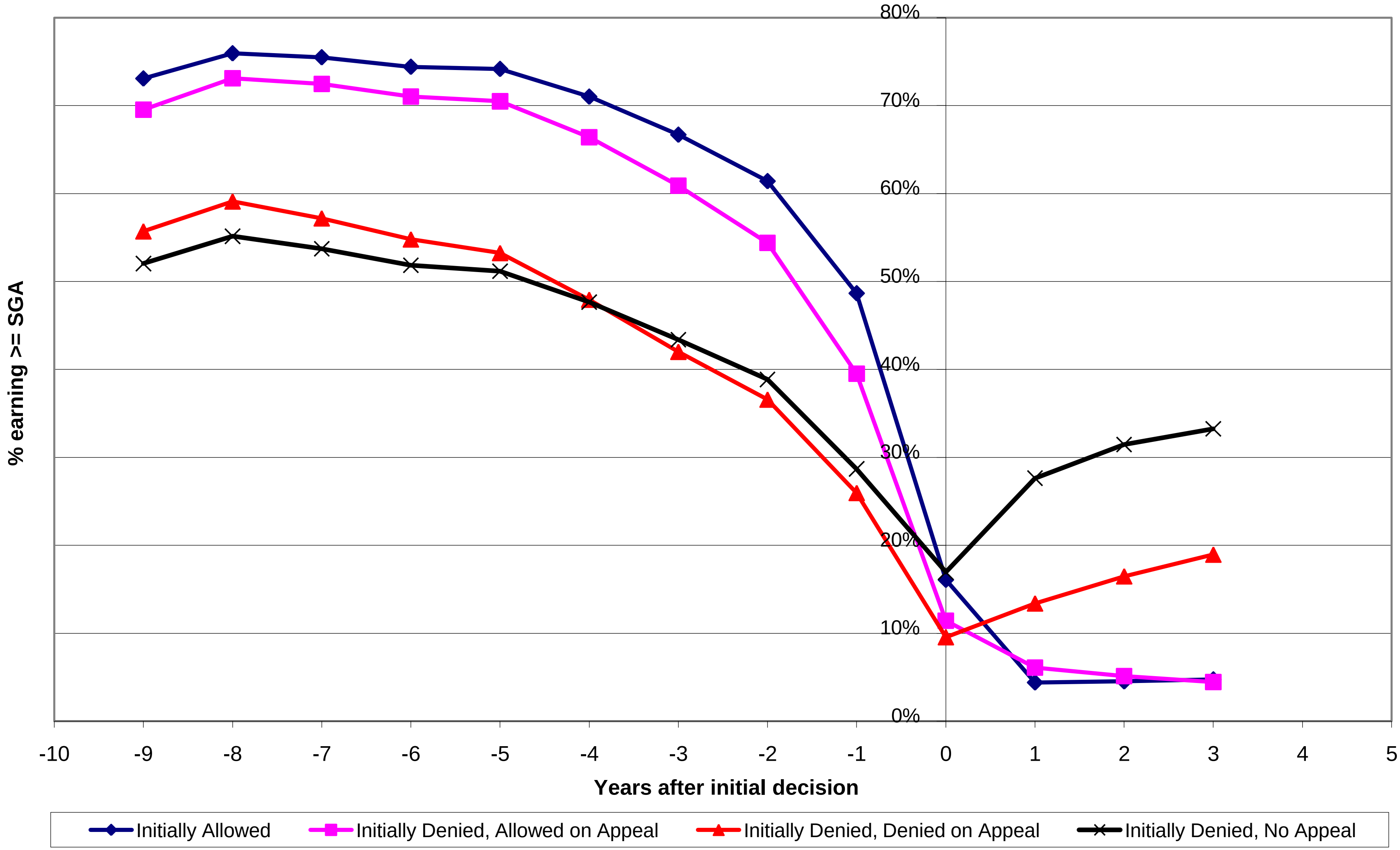


Figure 3C. Mean Earnings Before and After Initial Decision

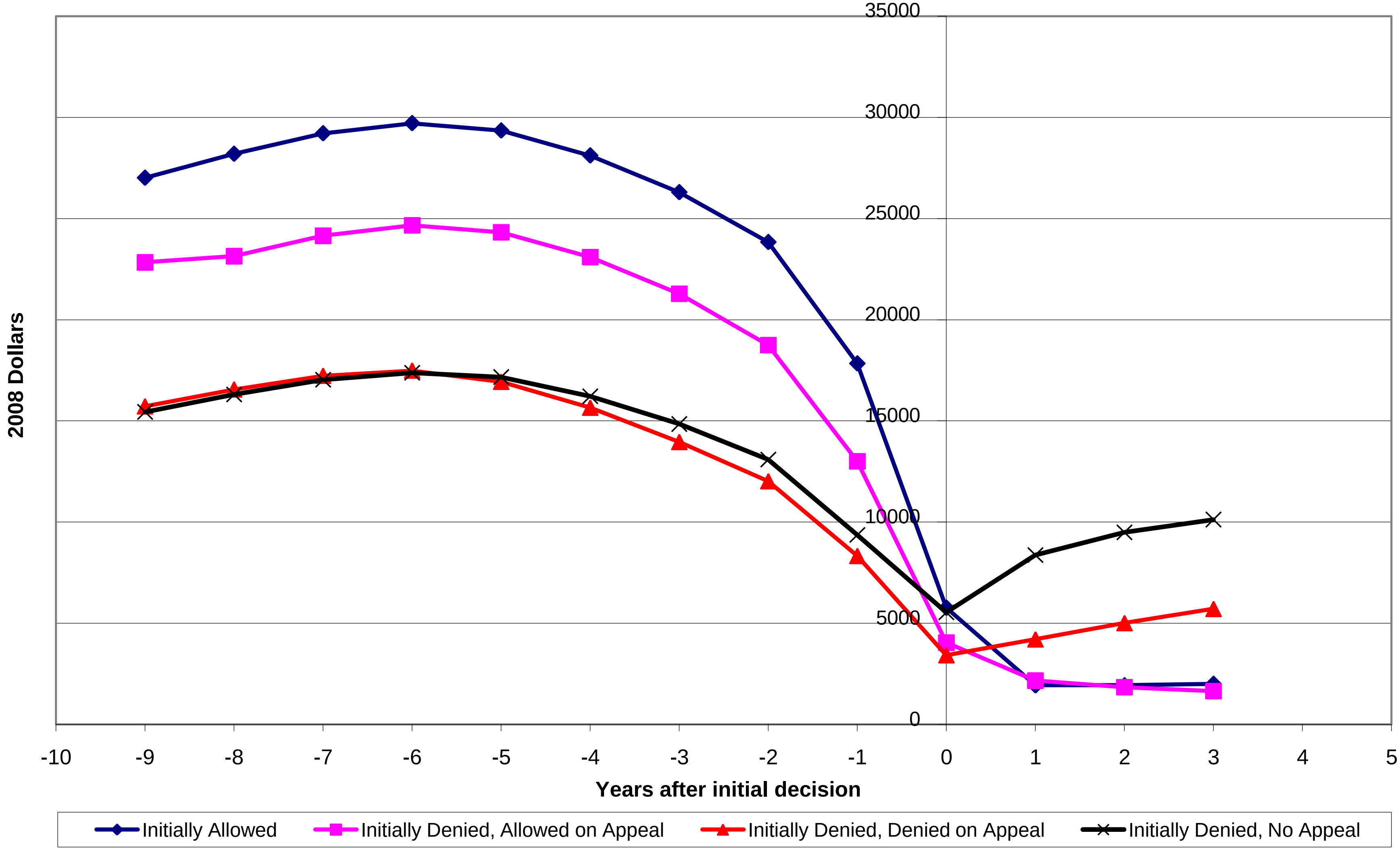
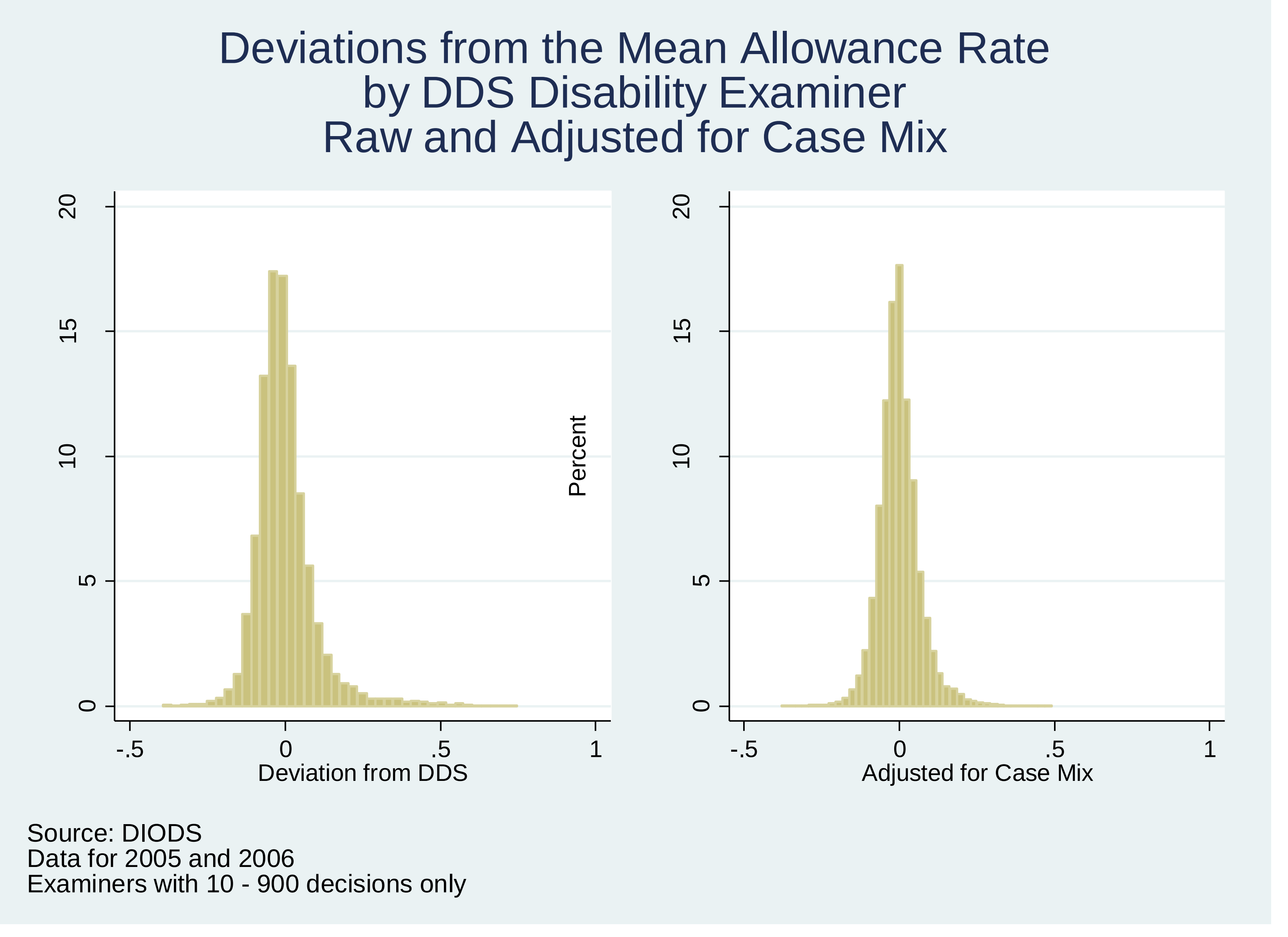


Figure 4



Deviations from the Mean Allowance Rate
by DDS Disability Examiner
Raw and Adjusted for Case Mix
Percent
0
5
10
15
20
-.5
0
.5
1
Deviation from DDS
Adjusted for Case Mix
Source: DIODS
Data for 2005 and 2006
Examiners with 10 - 900 decisions only

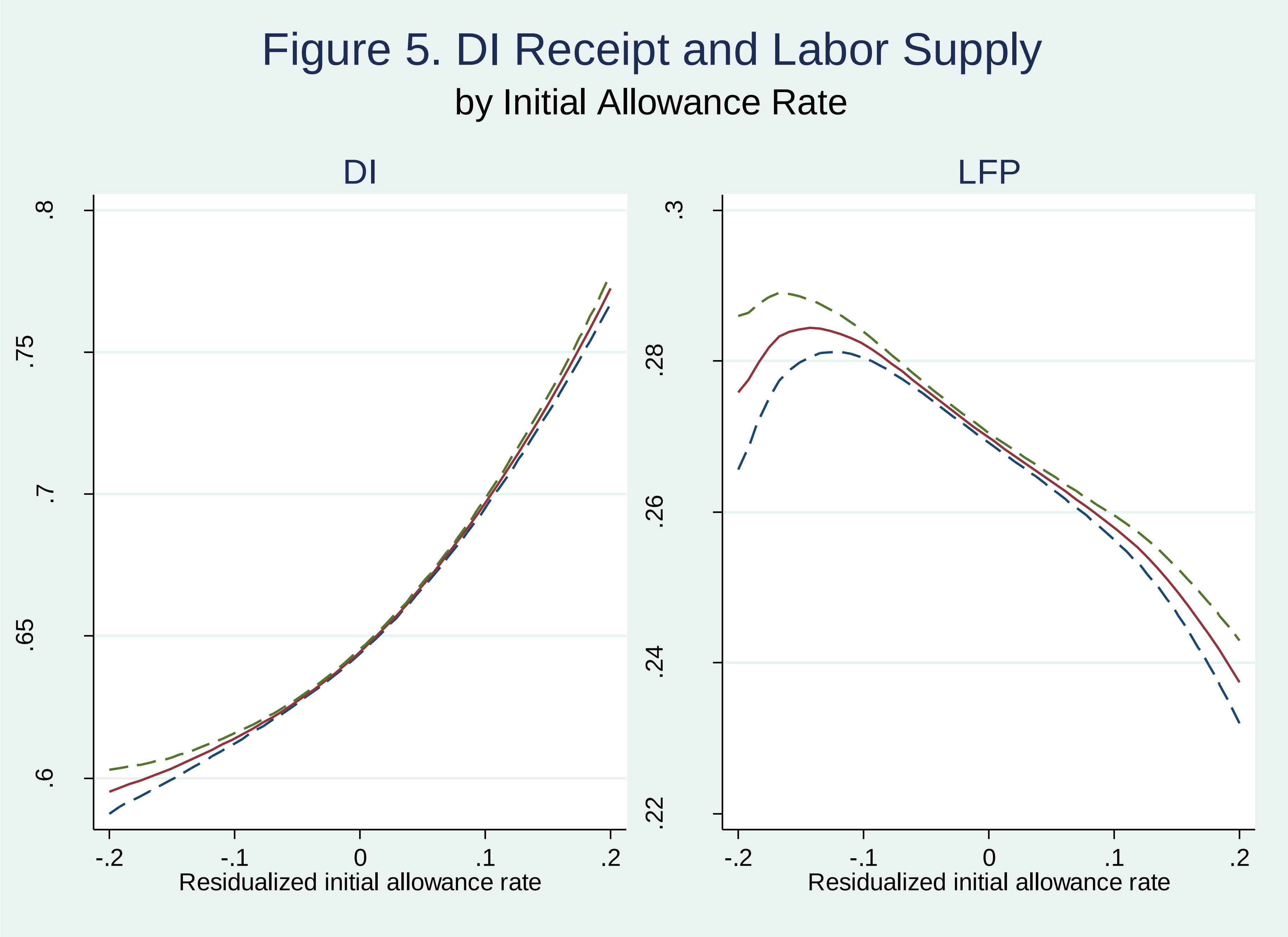


Figure 5. DI Receipt and Labor Supply
by Initial Allowance Rate
DI
LFP
.8
.75
.7
.65
.6
.3
.28
.26
.24
.22
-.2
-.1
0
.1
.2
Residualized initial allowance rate
Residualized initial allowance rate

Figure 6. Effect on Labor Force Participation by Age Group

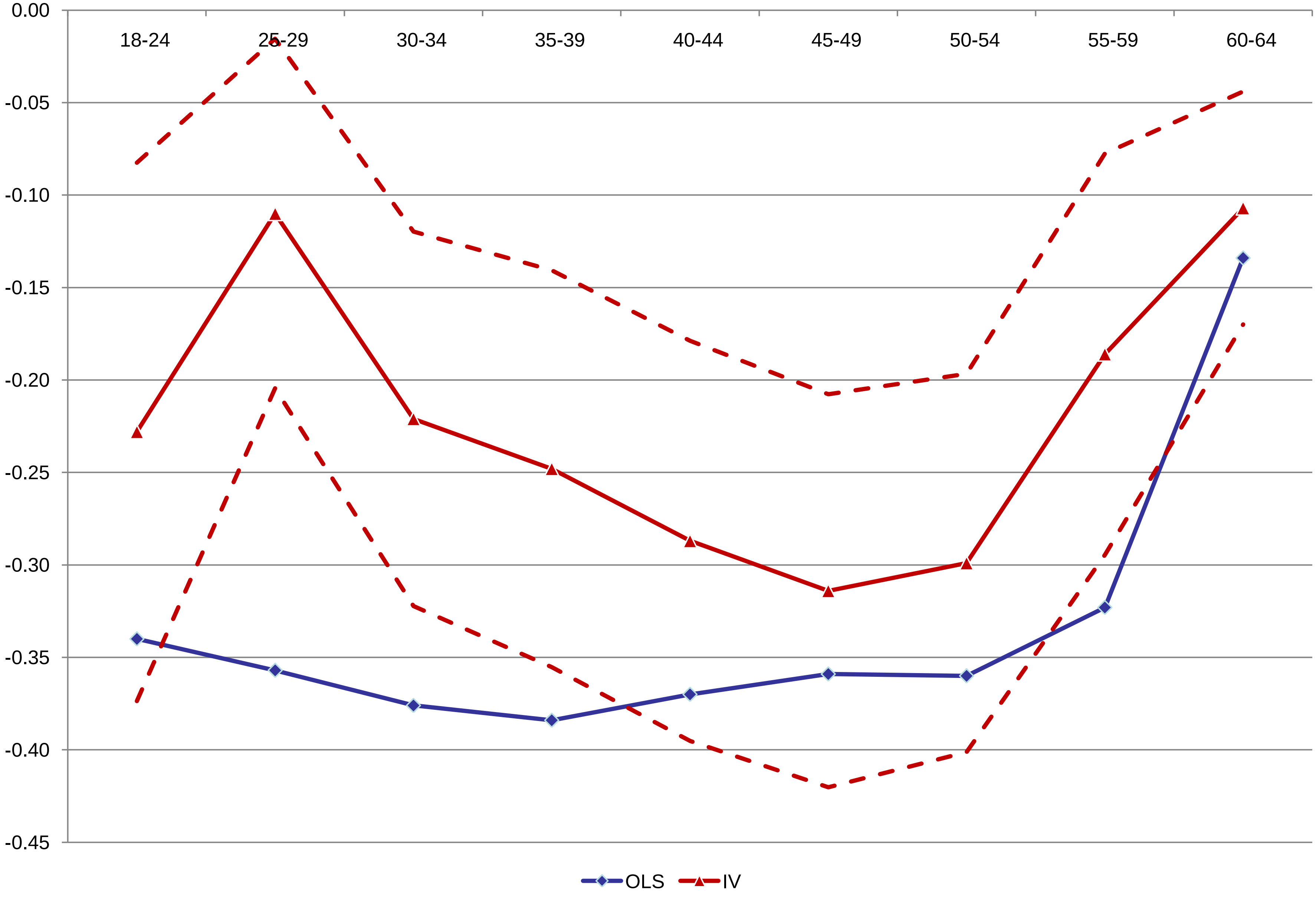


Figure 7. Marginal Treatment Effect of SSDI on Labor Force Participation