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

Nicole Maestas, Kathleen Mullen and Alexander Strand

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

May 2011

Michigan Retirement Research Center
University of Michigan
P.O. Box 1248
Ann Arbor, MI 48104
http://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 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.

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 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. 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 4 years from the date of the initial determination. By waiting two or more 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 could be changed because they were assigned to a particular examiner (Angrist and Pischke, 2009). We estimate that almost a quarter of applicants are on the margin of allowance. 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 (i.e., if their application had been denied). His or her likelihood of engaging in *substantial* gainful activity (SGA), defined by the SSDI program as earning more than $980 per month in 2009, 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

---

[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. This was pointed out by Parsons (1991).

participation of older beneficiaries would be at most 20 percentage points higher in the absence 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 marginal entrants under the present day system.

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 apply for SSDI benefits at their local field office, which screens out those who are not currently insured or are engaging in SGA. These are included in what is labeled "technical denials," and do not receive a medical review. The remaining applications are forwarded to a state Disability Determination Services (DDS) office, where cases are assigned to

disability examiners for review. We discuss assignment of applications to examiners in greater detail in the Section 3.1 and in the appendix.

Disability examiners 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 (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 impairments severe enough to prevent them from performing any SGA in the national economy given their vocational background). Vocational allowances are made in consultation with a medical-vocational grid which provides guidance for allowance decisions based on residual functional capacity, age group, education and type of 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). 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. Denials are issued to applicants engaging in SGA (step 1), impairments that are obviously temporary and non-severe (step 2), and if the individual has the residual functional capacity to perform any of their past jobs (step 4) or any work in the current economy (step 5).

Denied applicants can appeal their initial determination within 60 days by applying for reconsideration by the original DDS office.[2] 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 an SSA Appeals Council (AC) and finally to Federal Court. In our data, 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

---

[2] 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.

Inspector General (2008) found that in 2006 the average processing time for cases in the initial determination phase 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 for a period of five months before they are entitled to receive benefits, or until the allowance decision is made (whichever comes later). Up to 19 months of back payments are available depending on the onset and allowance decision dates. The average monthly benefit was $1,120 in 2009 (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. Finally, 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 leave the rolls for a reason other than death (35% in 2009) or automatic conversion to retired worker benefits (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 (Continuing Disability Review, or CDR) whereas 61% were found to earn more than SGA (Social Security Administration, 2010, Table 50).

3. Empirical Strategy

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

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

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 observed to be a SSDI beneficiary at any time after the initial determination), 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 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) 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.

From SSA's point of view, 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 impairment 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

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

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

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

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

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

where F_{s+v} denotes the cumulative distribution function for unobserved severity plus idiosyncratic examiner error, v_{ij}, which is assumed to be i.i.d. 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 implement an instrumental variables (IV) estimation strategy where we instrument for SSDI receipt in Equation 1 using assigned examiner's propensity P_j.[3]

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:

$$EXALLOW_{ij} = \frac{n\_allowed_j - 1(DI_i = 1)}{n\_cases_j - 1}.[4]$$

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.

---

[3] 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.

[4] 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 examiner allowance propensity as a continuous instrumental variable has the advantage of allowing us to compute marginal treatment effects, as can be seen below.

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-level initial allowance rate, and initial determinations may differ from final determinations as a result of appeals.

3.1. Conditional Random Assignment and Monotonicity Assumptions

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 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.

This assumption is plausible. Interviews with managers of twelve of the largest DDS offices revealed that nearly all cases are automatically assigned to examiners by a computer program that assigns new applications to examiners based on availability.[5] Managers at the DDSs variously described this process as a "round robin," "next in line," "rotate to the back," or "equal distribution" system. Our appendix describes the assignment based on this computer program in more detail.

An exception to the availability-based assignment process is that a small percent of cases are sent by the field offices with high-priority flags for expedited handling, and these are sometimes manually assigned to examiners that tend to specialize in high-priority cases. The most common high-priority case during our study period is an allegation of terminal illness, and these cases are almost always allowed in the initial determination. Nonetheless, we estimate that this potentially non-random assignment occurred for less than 0.5 percent of applications in our sample simply because these applicants do not survive through the end of the second year after their initial determination (i.e., long enough for us to measure post-decision labor supply outcomes). In addition, since terminal illness cases are heavily concentrated in certain

---

[5] Even prior to the age of computer assignment, there is evidence that cases were essentially 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.

observable diagnosis codes,[6] controlling for diagnosis code effectively controls for much of this non-randomness. [7]

A second assumption that is critical to our interpretation of our IV estimates is the assumption that examiners' award propensities affect applicants' chances of DI 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 more likely to be overturned on appeal than denials by lenient examiners, and thus the instrument is likely to have a weaker effect on individuals assigned to strict examiners. Below we provide evidence that the strength of the first stage increases with examiner leniency. A second implication of the monotonicity assumption is that examiners who are "strict" on one type of case, say musculoskeletal, are also strict on any other type of case, say mental disorders. We test this implication below, and find evidence that supports the assumption of monotonicity.

3.2. Heterogeneous Treatment Effects

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 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 (γ_i) is averaged over "compliers," or applicants whose treatment status could be changed as a result of assignment to a particular examiner. On the other hand, our instrumental variable is not informative for "always takers" – those who would be allowed, whether initially or upon appeal, even if assigned to the strictest 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.

---

[6] In particular, eight types of cancers and some other conditions are automatically classified as Terminal Illness cases, but other diagnoses may also receive this classification if the condition is untreatable.
[7] 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.

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 all 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).[8] 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

---

[8] 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.

(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 expected earnings difference between two otherwise identical individuals, one of whom is disabled while the other is not, 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 an increasing function of s_i; in other words, the magnitude of the disincentive effect is decreasing 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."[9] The DIODS contains alphanumeric codes that identify disability examiners uniquely within DDS and 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

---

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

cases in the observed time frame.[10,11] 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.

To examine past and subsequent labor supply, we link the applications data 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 2-4 years after the initial decision.[12] 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 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 in which earnings are observed.

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 (Social Security Administration, 2010, Table 59). The remaining 1.5 million applications in a year were sent to a DDS for a medical/vocational determination, and these appear in our data set. After imposing the sample restrictions discussed above and removing duplicate claims, we obtain a sample of just under 2.4 million observations for 2005 and 2006 combined. These cases were adjudicated by 7,193 DDS examiners.

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.[13] 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 –

---

[10] Unfortunately, we cannot distinguish between full- and part-time examiners, nor can we observe examiner tenure.
[11] 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 eliminate the 25 examiner codes with the largest numbers of decisions in order to avoid this potential problem.
[12] We observe earnings up to three years after the initial decision for the full sample, and up to four years later for 2005 applicants.
[13] 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.

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.

Figures 1-3 provide 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;[14] and 2) engaging in SGA (where the SGA threshold is that defined by SSA in a given year).[15] We also present mean earnings before and after the initial decision for each group. All earnings amounts are presented in 2008 dollars, and include zeros. From Figure 1, 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.

---

[14] 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.
[15] 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 per month in the 1990s and rose to $700 per month in 1999.

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.[16] 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 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 or fourth 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 (e.g., earning more than $11,760 per year in 2009) (see Figure 2) and average earnings (Figure 3). Interestingly, Figure 3 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 1 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

---

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

it would not be able to correct for differences in *unobservable* factors such as severity of the impairment, which 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

Although consistent treatment of similarly situated individuals is an important principle in any evaluative system, some degree of variation in outcomes is inevitable. In other settings, variation in outcomes has been associated with case complexity, dynamism of the subject matter, size of caseload, resources to assist decision-making, attributes of the evaluators such as their prior experiences, and training and guidance provided to evaluators (Ramji-Nogales, Schoenholtz and Schrag, 2007; Legomsky, 2007). Indeed, DDS examiners are called on to evaluate and weigh many aspects of complex cases against extensive medical and vocational criteria in a dynamic medical environment. They have heavy caseloads and are not trained physicians (although they may consult with physicians). These factors point to the importance of examiner judgment. 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 research design takes advantage of this natural and *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. In our sample, 90% of examiners are observed to handle more than 29 initial applications, and the median number is 215. 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 the 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. After adjustment, 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 separately plot smoothed SSDI receipt and labor force participation rates by examiner's residualized (regression-adjusted) initial allowance rate (with dashed lines indicating 95% confidence interval bands) estimated via local quadratic regression. SSDI receipt is increasing in the residualized initial allowance rate, but its slope is less than one on account of the appeals process. Its slope is also increasing in the residualized initial allowance rate since individuals initially denied by strict examiners are more likely to ultimately obtain benefits on appeal than individuals initially denied by lenient examiners. This pattern is consistent with monotonicity.

While SSDI receipt rises with the examiner's allowance propensity, the labor force participation rate of applicants two years after the initial decision falls, 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 which SSDI receipt increases; where the relationship between SSDI receipt and the initial allowance rate is relatively flat, labor force participation is also flat and only begins to decrease as DI receipt increases.

Table 2 presents the estimated coefficient on EXALLOW in a regression of individual allowance decisions, with and without covariates. We present results for 2005 and 2006, separately as well as pooled. 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 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 (our "leave oneself out" measure of allowance propensity) will affect the estimated coefficient on EXALLOW when included. Adding a finer

measure of geography by grouping individuals by the first three digits of their zip code (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 three-digit zip code dummies increased computing time substantially, we dropped them for the remaining analysis.[17]

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 in at least some DDS offices. This may be deliberate (e.g., an examiner tends to focus on high-priority cancer cases) or indirect (e.g., if certain types of cases 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.[18] Finally, controlling for age, average previous earnings and seasonality has very little impact on the estimated effect of the instrument on individual allowance decisions.[19] This suggests that, once one conditions on impairment type, applicants are approximately randomly assigned to examiners.

Conditional on observables, examiner allowance rates are strongly correlated with individual allowance decisions: 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 individuals' allowance decisions.[20] Note that in the case of a binary treatment, the first stage estimate also gives an estimate of the proportion of the sample who are compliers.[21] Thus, we estimate that approximately 23% of applicants represent "marginal" cases whose disposition depends on examiner assignment. By contrast, 77% of applicants would either be allowed or denied regardless of examiner assignment.

To test the assumption that examiners' generic allowance rates can be used to instrument for allowances across all types of cases, we performed the following test. First, we computed

---

[17] Since it is plausible that three-digit zip code 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 zip codes included and found that they did not affect those estimates either.

[18] Impairment type is incompletely measured where there are multiple diagnoses.

[19] It would also be useful to examine the effect of examiner characteristics (such as job tenure or holding Single Decision Maker authority) on examiner allowance rates, but we do not observe this information in our data.

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

[21] In our case, the proportion of compliers equals Pr(DI=1|EXALLOW=1)-Pr(DI=1|EXALLOW=0).

examiner allowance rates across all types of cases *except* cases with the same body system code. Then we regressed SSDI receipt for cases of the same body system on this new "leave body system out" measure of initial allowance rate. For example, we examine whether SSDI receipt for musculoskeletal cases is predicted by examiners' propensities to allow all other types of cases (mental, cardiovascular, etc.). Table 3 presents these estimates. While the strength of the instrument varies across impairment types, it is positive and statistically significant in all cases.

This body of evidence suggests that examiner allowance rates can be used effectively as an instrumental variable for individual allowance decisions.

6. Results

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

Table 4 presents results for the labor supply regressions estimated by OLS and IV. We model three outcomes: labor force participation (defined as earning more than $1,000/year), engagement in SGA (i.e., earning more than $11,760 per year in 2009), and annual earnings. We measure all labor supply outcomes two and three years after the initial decision, for applicants whose initial determination took place in 2005 or 2006, and we also examine outcomes four years later for those receiving an initial determination in 2005. In all models, we control for a rich set of covariates that include geography (DDS codes), characteristics of the impairment (diagnosis codes), age, prior labor market history, and seasonality (month of application).

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. 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.

Table 4 shows that despite their impairments, 14% to 15% of SSDI beneficiaries participated in the labor market – earning more than $1,000 in a given year – two years after their initial determination. Among denied applicants, roughly one-half worked two years after their initial determination. Adjusting for differences in observable characteristics such as impairment type, age or prior labor market history yields an estimated effect of SSDI receipt on

labor force participation of approximately 31-34 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 and Work. In other words, no more than half of SSDI beneficiaries would be likely to engage in any labor market activity 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 between the two groups – yields smaller coefficient estimates, between 21 and 22 percentage points after two years.

Note that while estimates of the effect after two years are similar for applicants with 2005 and 2006 determinations, estimates of the effect after three years are divergent: 17 percentage points for individuals with decisions received in 2005 (and thus labor force participation measured in 2008) vs. 14 percentage points for individuals with decisions received in 2006 (labor force participation measured in 2009). Furthermore, Figure 6 illustrates that the estimated coefficients display an unexpected pattern where they decline in magnitude as time since the initial determination increases. This pattern is not due to the phenomenon of denied applicants strategically depressing labor supply while waiting for a decision on appeal, since as appealed cases resolve over time and some denied appellants return to the labor market, our estimates of the labor supply effects of SSDI receipt should *increase* in magnitude over time rather than decrease.

A plausible explanation for these patterns is that the estimated effect of SSDI receipt on labor force participation reflects both the supply of *and* demand for disabled workers, and that as the financial crisis deepened and the economy worsened between 2007 and 2009, it became harder for denied applicants to find employment, even if they were willing and able to work. In the time period over which we observe labor supply, the national unemployment rate skyrocketed from roughly 5% to more than 9.5% between the beginning and end of 2008; in contrast, note that the unemployment rate fluctuated relatively steadily between 4% and 6% between 2001 and the beginning of 2008. As a result, we focus on our two-year estimates of approximately 21 percentage points as representative of the effect of SSDI receipt on labor force participation under "average" economic conditions. This facilitates comparison of our estimate with previous estimates in the literature that were generated under similar economic conditions.

To more fully understand the magnitude of our estimates it is useful to translate our percentage point effect into a percent effect by dividing by the appropriate baseline labor force participation rate. Because we estimate the average effect of SSDI receipt over the population of "marginal entrants," this is not straightforward. In particular, because the marginal entrant is likely to be less healthy than the average denied applicant, then the labor force participation rate of all denied applicants is likely to be an *upper* bound for the labor force participation of the marginal entrant in the absence of SSDI. As a result, the estimated labor supply effect as a percent of average labor force participation among denied applicants represents a *lower* bound on the magnitude of the effect. Similarly, comparing the point estimate to the labor supply of beneficiaries will give an upper bound on the magnitude of the effect.[22] Thus, we find that the estimated effect of SSDI receipt among marginal applicants is large – between 42% and 140% of baseline labor force participation.

As noted earlier, SSA suspends SSDI benefits for individuals earning more than the SGA threshold, which in 2009 was set at $980 per month, or $11,760 per year. 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 4 shows that 4-5% of beneficiaries earned more than the annualized SGA threshold in a given year.[23] Somewhat surprisingly, however, less than 30% of denied applicants were performing substantial gainful activity up to four years after their initial determination. IV estimates of the effect of SSDI receipt on earning more than the SGA threshold are 11-13 percentage points – again, relatively large in percent terms, between 45% and 250% for the marginal entrant.[24]

---

[22] 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.

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

[24] Note that since we are using administrative earnings records, it is possible that we are undercounting earnings that are "under the table," or not reported to the government. Since this underreporting is more likely to be prevalent among SSDI beneficiaries whose earnings are restricted, this would tend to bias our estimated effect away from zero. By the same token, we may be overstating the upper bound on the magnitude of the effect.

Finally, we examine average earnings losses associated with SSDI receipt. Table 4 shows that mean annual earnings of SSDI beneficiaries are between approximately $1,500 and $2,000 per year, with higher earnings losses in earlier calendar years. Denied applicants earn between $7,200 and $8,900 per year on average – below the 2009 annualized SGA threshold of $11,760. IV estimates range from a high of approximately $2,600 two years later for 2006 determinations (compared with $1,600 two years later for 2005 determinations) to a low of a statistically insignificant $375 four years after the (2005) initial determination for earnings measured in 2009. Again, the economy is seen to play an important role in determining the labor supply effects of SSDI. Our estimates suggest the marginal denied applicant is worse off than the allowed applicant in terms of total income: while the denied applicant does not receive an average benefit of roughly $12,000 annually, he only increases his earned income by $1,600-$2,600.

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 5 and 6 present first stage and second stage results, respectively, for these disaggregated groups. As shown in Table 5, 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. Angrist and Pischke (2009) show how one can 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).[25]

Specifically, the relative likelihood that the marginal entrant has a particular observable characteristic is given by the ratio of the first stage coefficient conditional on that characteristic relative to the overall first stage coefficient. Thus, a policy change that uniformly increased the allowance threshold at the initial determination stage would induce onto SSDI individuals who were 50% more likely than the average applicant to suffer from a mental disorder, and 30% less likely than average to suffer from a musculoskeletal impairment. This is not surprising, given

---

[25] Note that the monotonicity assumption must be satisfied for this result to hold.

that 64% of allowances for mental disorders result from initial determinations, compared with 38% of allowances for musculoskeletal impairments.[26] Similarly, a policy that increases the allowance threshold 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 6 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 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 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. 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

---

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

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 as a function of predicted SSDI receipt in the range of 55 percent to 75 percent (with 95% confidence intervals), which encompasses the ultimate 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.

Note that our interpretation of the MTE as a function of unobserved severity depends critically on two assumptions: (1) the monotonicity assumption already described, and (2) the assumption that assignment to initial examiner *only* effectively randomizes SSDI receipt and not some other outcome that independently affects labor supply outcomes. For example, if strict examiners tend to process claims more slowly and/or send more applicants into the appeals process, then applicants on the margin for strict examiners will tend to spend more time out of the labor market while awaiting appeal. *If* time out of the labor market adversely affects employment outcomes (e.g., through signaling to an employer or by depreciating human capital) then the MTE will tend to *understate* the estimated effect of SSDI receipt on labor force participation for lower severity applicants (because it also encompasses the effect from longer waiting times). Figure 8 presents average initial and final processing times (in days) by predicted SSDI receipt given the initial examiner. Note that both very strict and very lenient examiners tend to process claims more slowly than the average examiner, although lenient examiners are even slower than strict examiners. However, since strict examiners tend to generate more appeals than lenient examiners, on average applicants assigned to very strict or very lenient examiners tend to have similar final processing times. In any case, the U-shaped pattern in processing times by examiner allowance propensity implies that the declining pattern in the MTE (corresponding to a disincentive effect that rises in severity) cannot solely be driven by differences in processing times.

Our finding that the MTE is declining in unobserved severity may explain 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 subsample of applicants who appeal their initial determination likely consists of higher-severity individuals on average (i.e., individuals with higher expected returns to appeals);

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.[27] On the other hand, the Chen-van der Klaauw estimate is based on 55 year olds drawn from the middle of the severity distribution (on the margin of vocational allowances) 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 rise—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. Ours are 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 labor force participation of the marginal entrant would be on average 21 percentage points greater in the absence of SSDI benefit receipt under "average" economic conditions. 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

---

[27] Although not directly comparable to our estimates, French and Taber (2010) cite estimates of a declining MTE of SSDI receipt on earnings, ranging from -$1,500 to -$3,500, based on the data of French and Song (2009).

26

$1,600 to $2,600 more per year on average in the absence of SSDI benefit receipt. We estimate that in the present system, nearly one-quarter of applicants are "marginal" cases whose ultimate outcome depends on 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 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 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.

References

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

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 (2007). "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review*, 97(5), pp. 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), pp. 746-770.

Foote, S. M. and C. Hogan (2001). "Disability Profile And Health Care Costs Of Medicare Beneficiaries Under Age Sixty-Five." <u>Health Affairs,</u> 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.

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, 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.

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

Legomsky, Stephen H. (2007). "Learning to Live with Unequal Justice: Asylum and the Limits to Consistency." *Stanford Law Review*, 60:413.

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.

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

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

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

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

Ramji-Nogales, Jaya, Andrew I. Schoenholtz, and Philip G. Schrag (2007). "Refugee Roulette:
Disparities in Asylum Adjudication. *Stanford Law Review*, 60: 295.

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.

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), pp. 518-529.

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."Forthcoming, *American Economic Review*.

Table 1. Summary Statistics

Variable	All Applicants	Initially Allowed	Initially Denied		
			Allowed on Appeal	Denied on Appeal	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. Rejected - nonsevere	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. Rejected - work past jobs	23.64%	0.00%	38.25%	34.31%	32.72%
5. Rejected - able to work	25.69%	0.00%	43.72%	39.25%	32.06%
5. Allowed - unable to 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					
Mean	22,100	27,918	22,895	15,514	16,078
Std. deviation	26,879	33,117	24,122	20,853	24,466
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. Alternative First Stage Estimates Using "Leave Body System Out" Instrument

Excluded impairment	No. obs.	Coeff. On EXALLOW2	t-Stat.
Musculoskeletal System	881,058	0.098	11.13
Special Senses and Speech	58,603	0.426	19.19
Respiratory System	93,941	0.138	7.17
Cardiovascular System	195,183	0.166	10.99
Digestive System	56,119	0.193	6.98
Genitourinary Impairments	34,835	0.093	5.81
Hematological Disorders	8,266	0.407	6.93
Skin Disorders	7,933	0.366	4.7
Endocrine System	98,801	0.065	2.78
Neurological	194,148	0.207	16.1
Mental Disorders	513,884	0.251	21.48
Malignant Neoplastic Diseases	80,251	0.234	12.5
Immune System Disorders	63,765	0.206	7.4
Special/other	93,457	0.077	3.58

Notes: Each row represents a regression of ultimate SSDI receipt on examiner allowance rate constructed using pooled 2005-06 cases except those with the same body system code. Control variables include: DDS dummies, diagnosis codes, age group dummies, avg. previous earnings and month-year dummies. Standard errors clustered on examiner. Note 11 cases excluded from musculoskeletal regression that belonged to a complete specializer.

Table 4. Effects of SSDI Receipt on Labor Force Participation & Earnings

| | 2005 Decisions | | | | | | 2006 Decisions | | | |
| | 2 Years After Decision | | 3 Years After Decision | | 4 Years After Decision | | 2 Years After Decision | | 3 Years After Decision | |
Outcome	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
1) Earn >=$1,000/year										
Mean dependent variable | allowed	0.154		0.136		0.114		0.139		0.112	
Mean dependent variable | denied	0.516		0.506		0.458		0.479		0.433	
Coeff. on ALLOW	-0.338***	-0.215***	-0.349***	-0.170***	-0.328***	-0.103***	-0.322***	-0.206***	-0.310***	-0.144***
	(-326.10)	(-6.75)	(-338.53)	(-5.42)	(-318.19)	(-3.34)	(-304.66)	(-8.58)	(-287.92)	(-6.12)
R-squared	0.197	0.183	0.211	0.180	0.197	0.143	0.188	0.174	0.181	0.151
2) Earn >= SGA										
Mean dependent variable | allowed	0.052		0.046		0.038		0.045		0.035	
Mean dependent variable | denied	0.286		0.291		0.257		0.256		0.229	
Coeff. on ALLOW	-0.235***	-0.129***	-0.244***	-0.109****	-0.218***	-0.066***	-0.215***	-0.108***	-0.197***	-0.088***
	(-261.97)	(-6.11)	(-267.38)	(-5.33)	(-251.29)	(-3.30)	(-227.61)	(-6.09)	(-212.58)	(-5.19)
R-squared	0.143	0.125	0.157	0.126	0.143	0.100	0.133	0.112	0.126	0.102
3) Earnings										
Mean dependent variable | allowed	2,012		1,834		1,628		1,775		1,498	
Mean dependent variable | denied	8,671		8,857		8,104		7,722		7,164	
Coeff. on ALLOW	-7,150***	-1,634*	-7,343***	-1,562**	-6,752***	-375	-6,529***	-2,564***	-6086***	-1508***
	(-140.81)	(-1.71)	(-193.30)	(-2.01)	(-182.34)	(-0.47)	(-92.92)	(-3.92)	(-126.87)	(-2.49)
R-squared	0.131	0.094	0.139	0.093	0.119	0.064	0.134	0.112	0.121	0.087
No. observations	1,174,852		1,152,585		1,123,846		1,205,403		1,173,926	

Notes: t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.10. ALLOW denotes ultimate award decision (rather than initial decision). Control variables include: DDS dummies, diagnosis codes, age group dummies, avg. previous earnings, and month dummies.

Table 5. Heterogeneity: First Stage Regressions, 2005 and 2006 Combined

|	No. Obs.	Initial Allowance Rate	Ultimate Allowance Rate	Coeff. on EX-ALLOW	t-Stat.	Relative Likeli-hood
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: *** 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 6. Heterogeneity: Effects of SSDI Receipt on Labor Force Participation (2 Years Later), 2005 and 2006 Combined

				OLS		IV	
	No. Obs.	Mean LFP | Allowed	Mean LFP | Denied	Coeff. On ALLOW	t-Stat.	Coeff. On ALLOW	t-Stat.
Body system code							
Musculoskeletal system	881,069	0.125	0.509	-0.349***	-303.13	-0.195***	-4.33
Special senses and speech	58,603	0.209	0.523	-0.295***	-69.27	-0.084*	-1.71
Respiratory system	93,941	0.099	0.457	-0.294***	-81.97	0.108	0.80
Cardiovascular system	195,183	0.116	0.451	-0.316***	-131.96	-0.332***	-4.43
Digestive system	56,119	0.157	0.516	-0.344***	-80.89	-0.310**	-2.46
Genitourinary impairments	34,835	0.192	0.532	-0.310***	-36.77	0.470	1.38
Hematological disorders	8,266	0.242	0.586	-0.322***	-28.43	-0.044	-0.27
Skin disorders	7,933	0.158	0.555	-0.374***	-34.70	-0.604***	-3.25
Endocrine system	98,801	0.127	0.460	-0.308***	-101.82	-0.396	-1.39
Neurological	194,148	0.124	0.530	-0.368***	-140.74	-0.263***	-4.60
Mental disorders	513,884	0.186	0.532	-0.317***	-204.78	-0.212***	-7.78
Malignant neoplastic diseases	80,251	0.211	0.607	-0.377***	-93.44	-0.316***	-4.67
Immune system disorders	63,765	0.174	0.521	-0.314***	-64.24	-0.163	-1.64
Special/other	93,457	0.133	0.318	-0.253***	-70.09	-0.636	-1.60
Age at decision							
18-24	78,946	0.363	0.698	-0.340***	-88.45	-0.228***	-3.07
25-29	136,461	0.273	0.610	-0.357***	-109.80	-0.110**	-2.28
30-34	156,838	0.232	0.586	-0.376***	-124.77	-0.221***	-4.28
35-39	211,452	0.205	0.566	-0.384***	-171.92	-0.248***	-4.53
40-44	295,526	0.174	0.523	-0.370***	-197.96	-0.287***	-5.20
45-49	370,632	0.148	0.490	-0.359***	-200.68	-0.314***	-5.79
50-54	399,274	0.127	0.466	-0.360***	-204.98	-0.299***	-5.73
55-59	413,497	0.102	0.402	-0.323***	-152.58	-0.186***	-3.36
60-64	317,629	0.096	0.217	-0.134***	-79.26	-0.107***	-3.33
Avg. prior earnings							
Bottom quintile	476,051	0.148	0.332	-0.181***	-141.19	-0.176***	-8.66
2nd quintile	476,051	0.164	0.530	-0.337***	-232.30	0.022	0.46
3rd quintile	476,051	0.149	0.582	-0.405***	-266.42	-0.062	-1.17
4th quintile	476,051	0.135	0.606	-0.444***	-277.59	-0.225***	-4.70
Top quintile	476,051	0.141	0.597	-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. Labor Force Participation Before and After Initial Decision
2005 Decisions
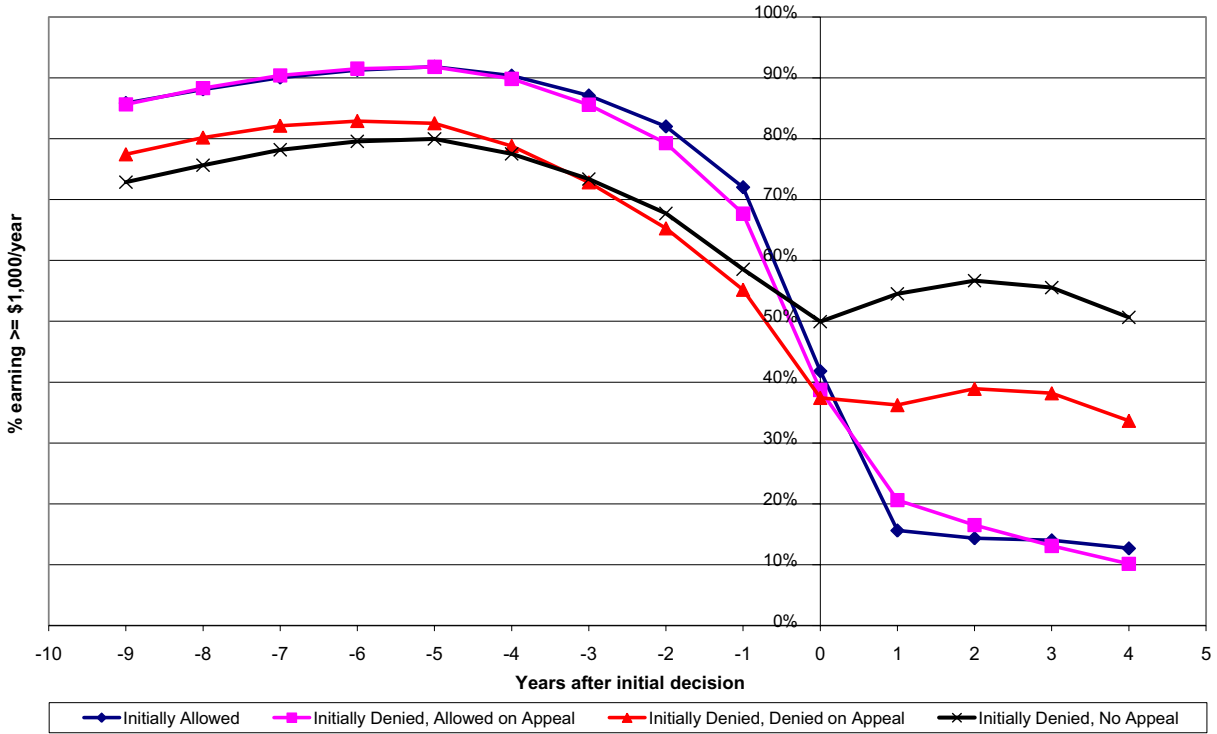
Figure 2. % Performing Substantial Gainful Activity (SGA) Before and After Initial Decision
2005 Decisions

36
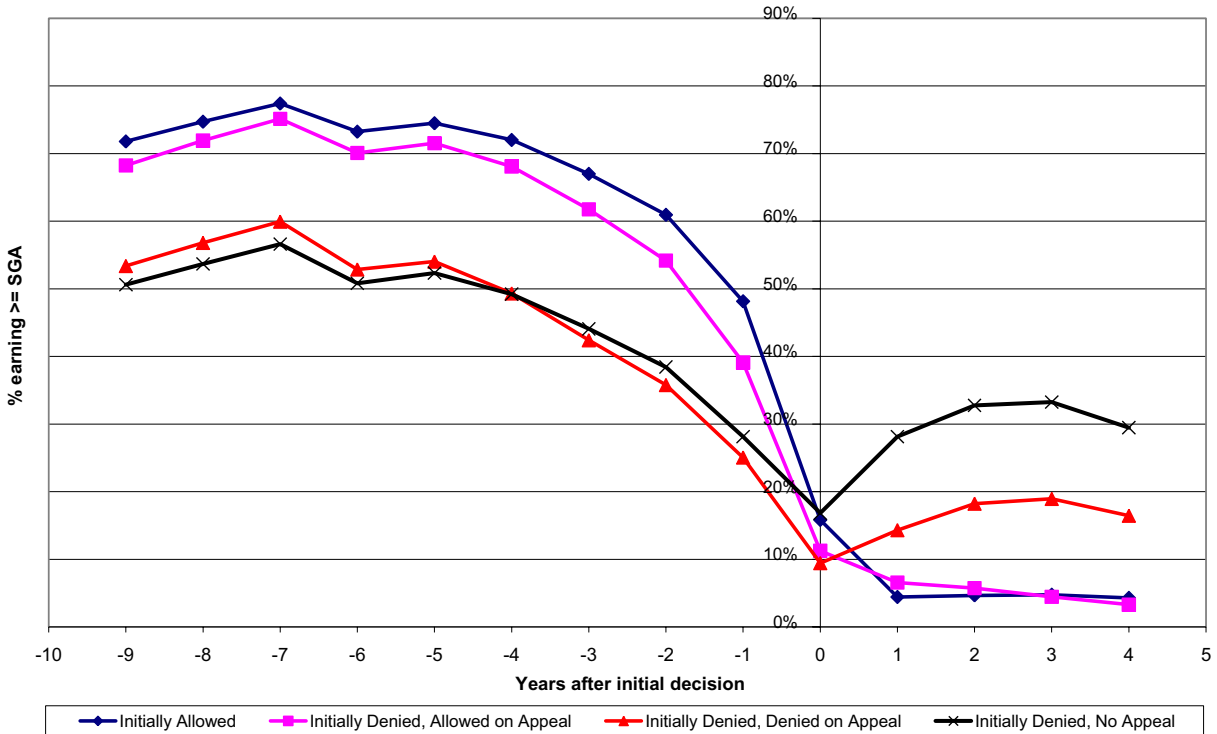
Figure 3. Mean Earnings Before and After Initial Decision
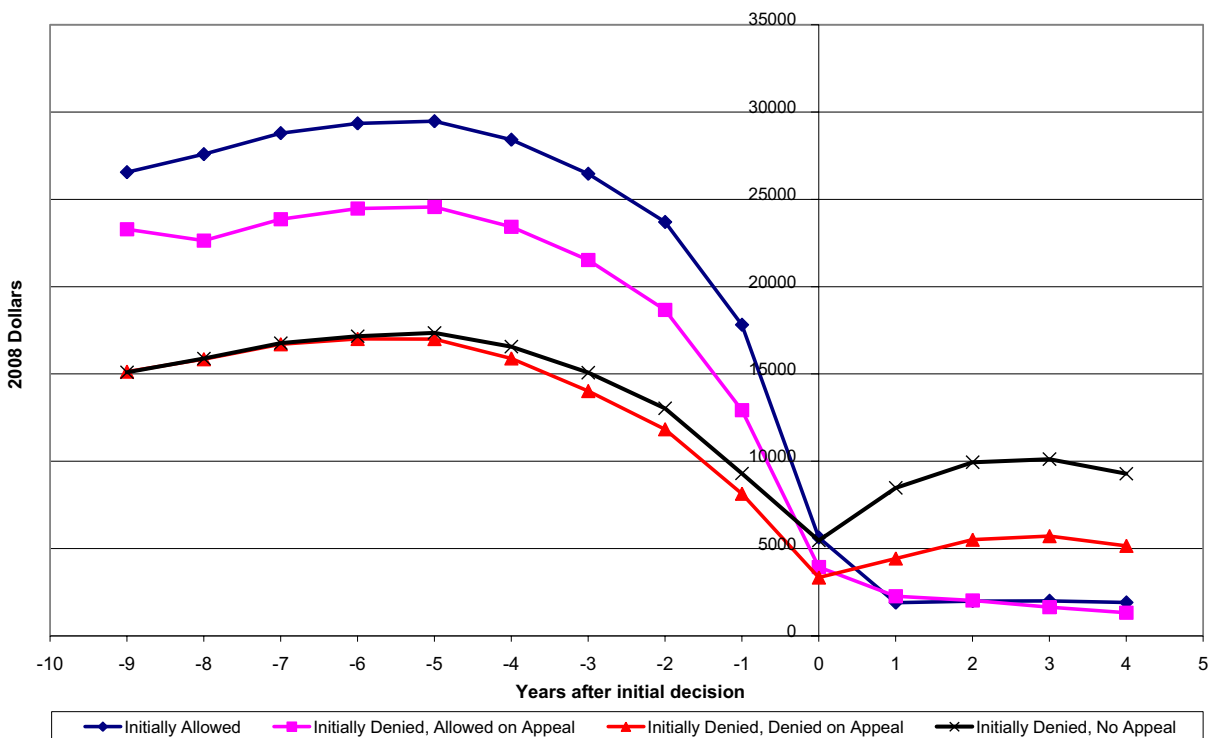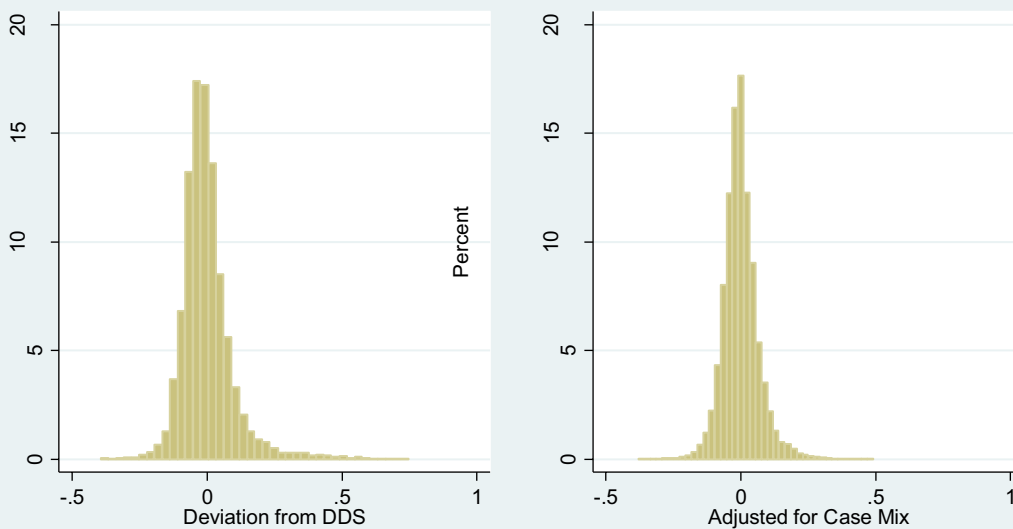**2005 Decisions**

Figure 4

Deviations from the Mean Allowance Rate
by DDS Disability Examiner
Raw and 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

Source: DIODS data for 2005 and 2006. Examiners with 10 - 900 decisions only.
Confidence intervals shown with dashed lines.
Labor force participation measured in the second year after the initial decision.

Figure 6. Estimated Effects of SSDI Receipt on Labor Force Participation

Note: Confidence intervals shown with dashed lines.

Figure 7

Estimated Marginal Treatment Effect

Source: DIODS data for 2005 and 2006. Examiners with 10 - 900 decisions only.
Confidence intervals shown with dashed lines.
Labor force participation measured in the second year after initial decision.

Figure 8

Initial and final processing times

Initial decision time   Final decision time

Source: DIODS data for 2005 and 2006. Examiners with 10 - 900 decisions only.
Confidence intervals shown with dashed lines.
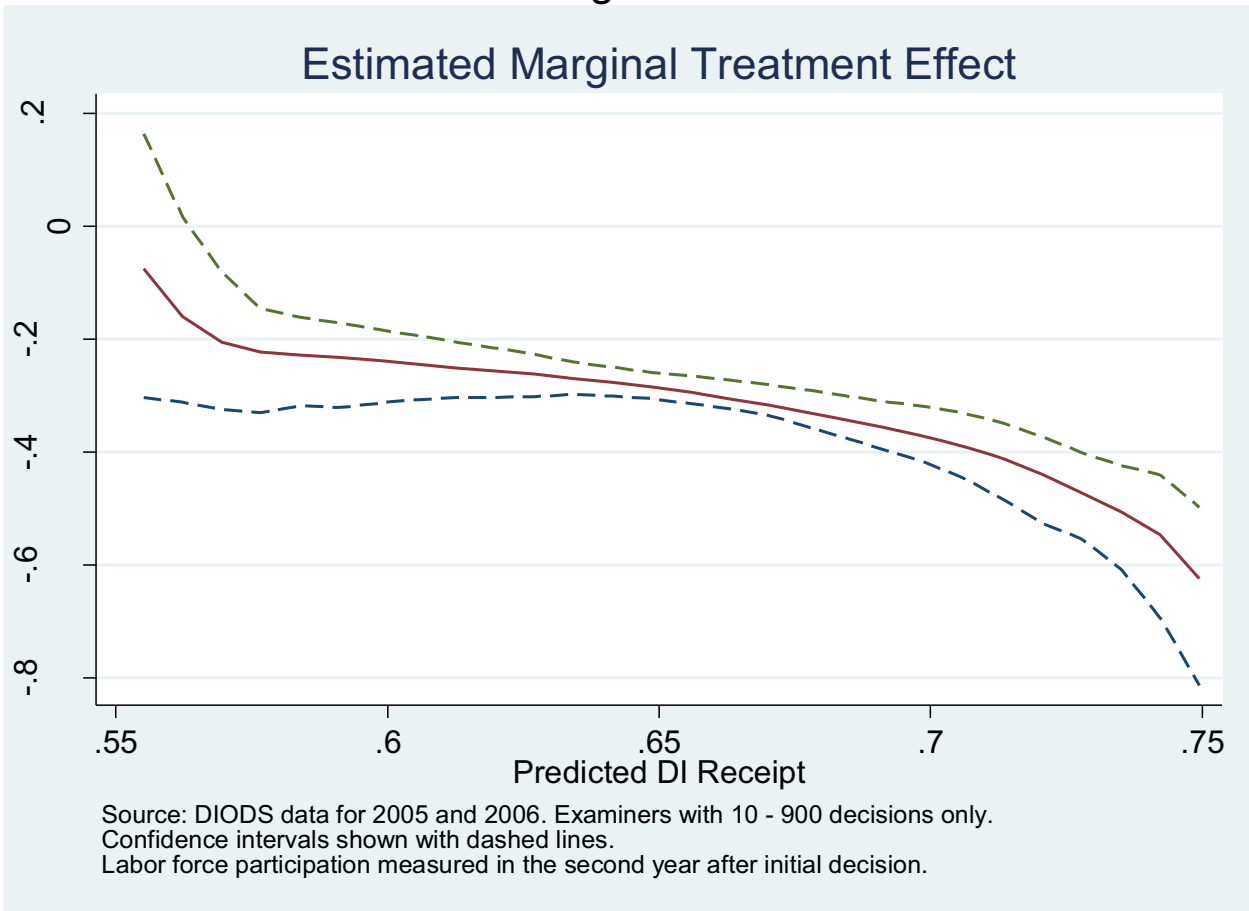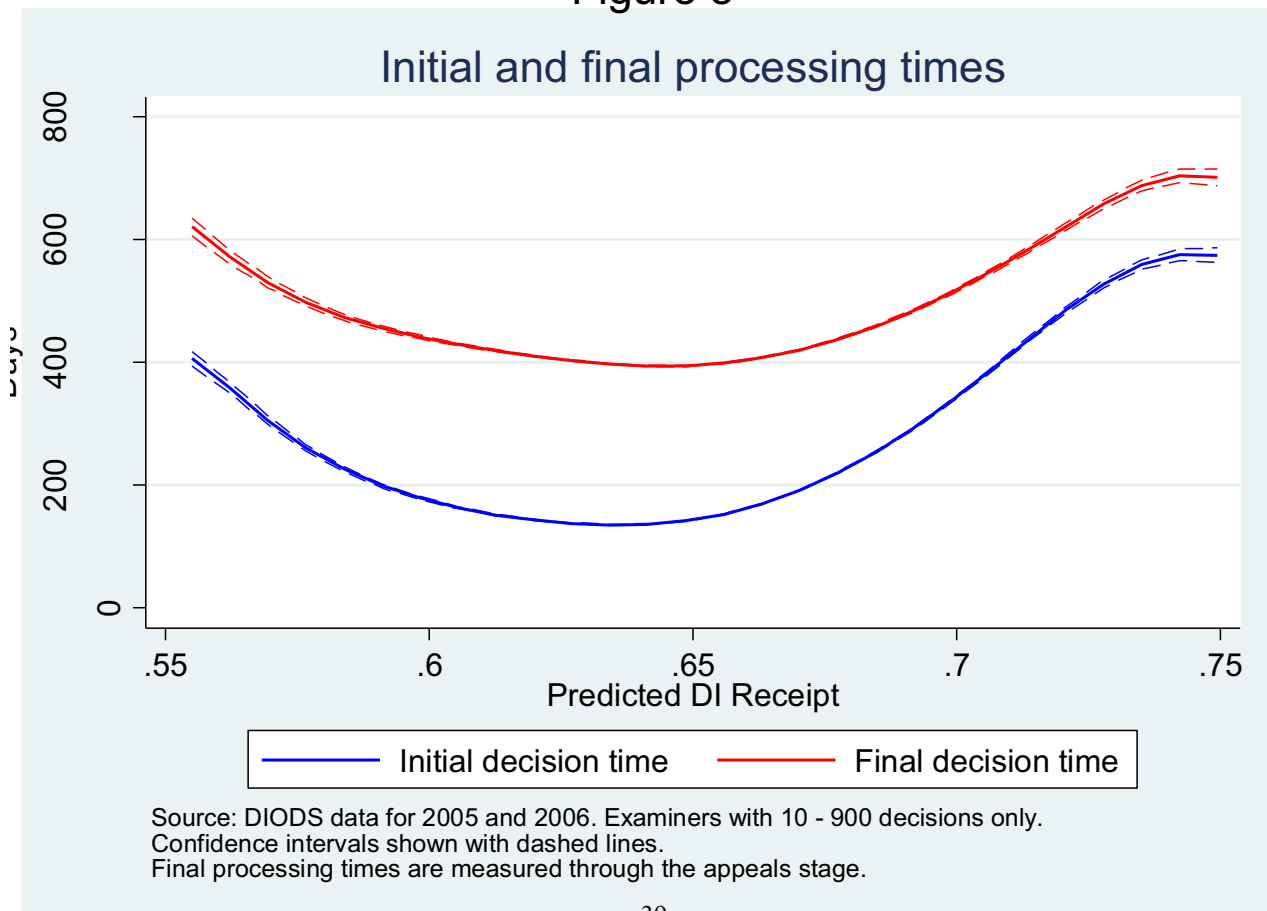Final processing times are measured through the appeals stage.

Appendix

Automated and Manual Allocation of Claims to Disability Examiners

This appendix describes the allocation of SSDI claims to disability examiners in 2005 and 2006 based on interviews with managers in the Disability Determination Services (DDSs) that had the most initial determinations in 2006. These twelve DDSs processed 40 percent of the claims in our sample in 2006.[1]

The vast majority of initial claims for adults were allocated to examiners using software developed by Iron Data LLC. While this software is currently still in use, SSA announced in January 2011 that it will switch to a new processing system that will be provided by Lockheed Martin with Iron Data LLC acting as a subcontractor.

The existing software allocates disability cases to examiners based on the availability of examiners. Managers at DDSs have described this process as a "round robin," "next in line," "rotate to the back," and "equal distribution" system. The case characteristics do not affect the distribution in any way, except in special circumstances described below. Allocations are done frequently – daily in many DDSs, continuously in others – and case characteristics are not known beforehand.

DDS staff utilize an electronic interface that for each specific allocation defines the availability of examiners for specific types of cases. There are generally three levels of examiner availability; they can be available for 1) initial claims only, 2) initial claims plus reconsiderations, or 3) initial claims and reconsiderations plus continuing disability reviews.[2] Thus, all examiners were available for initial adult cases during the analysis period. Although all examiners were not available with equal probabilities,[3] the probability of assignment to a particular examiner given availability did not vary across initial adult claims. The particular assignment of a claim to an examiner depended merely on the order of the receipt of claims and

---

[1] We intererviewed managers at the ten largest DDSs as well as in two states that are known to have used their own computer software to allocate cases during the analysis period, New York and Missouri.

[2] Names for these classifications of examiners vary among DDSs. Some DDSs have additional levels corresponding to child cases or other types of specializations. Trainees are also generally in a unique classification. Some DDSs allocated cases to trainees through the Iron Data system and some allocated them manually. The trainee period varied among DDSs; a range of 3 to 18 months was described by DDS managers.

[3] For example, the number of claims that could be assigned to levels 2 and 3 examiners can be constrained by the number of reconsiderations and continuing disability reviews that have already been assigned. Some managers indicated that these claims are assigned before initial claims are assigned. Also, trainees generally process fewer claims on a daily or weekly basis.

the order of the previous allocation. That is, an examiner was placed at the back of the queue of available examiners upon being allocated a claim.

The most common exception to computer allocation is for claims that are marked as high priority by the field offices and, thus, qualify for expedited handling. These claims are more quickly processed by a parallel team of examiners. In some DDSs, these examiners only work on expedited cases and, in some DDSs, the teams overlap. Further, some DDSs use the Iron Data system to allocate expedited claims and some do it manually. The types of expedited claims include:

- Quick Disability Determinations (QDD);
- Compassionate Allowances (CAL) ;
- Military Service Casualties ("Wounded warriors"); and
- Terminal Illness cases.

Expedited claims comprise a very small portion of our sample during 2005 and 2006. One reason is that QDD and CAL were begun after this period.[4] Another reason is that the number of Military Service Casualties was a very small percentage of the total.[5] Finally, Terminal Illness cases were more common but very few of these individuals survive to the second year after the initial determination, and therefore do not qualify for our sample.[6]

Another parallel system of allocation exists for claims for which the manager decides to allocate manually. In cases where there is a conflict of interest, the claims are commonly sent to another DDS for processing. In cases where the claim has a large amount of publicity or is the subject of a court case, the manager may assign a senior examiner. DDS managers indicated that these cases occur between less than once a month up to several times a month.

Some examiners have Single Decision Maker (SDM) authority, which allows them to make a determination without the use of medical consultants. While the extent and

---

[4] The national rollouts of QDD and CAL were in September 2007 and October 2008 respectively.

[5] An audit by SSA's Office of the Inspector General found 7,127 such cases in 2008, or less than 0.3 percent of determinations. See Social Security Administration, Office of the Inspector General, Military Service Casualty Cases, December 2009, A-01-09-29056.

[6] Terminal illness cases are characterized by allegations that the impairment cannot be reversed and is expected to end in death. According to SSA's management data, there were 44,904 Terminal Illness cases in 2007, or 3.5 percent of all initial claims. Further, an internal SSA study of 15,970 Terminal Illness allowances for 2005 found that the one-year survival rate from the time of allowance was around 11 percent. It follows that very few of these claims will be observed in our data because claimants who did not survive completely through the second year after the initial decision were dropped. In order to be observed in the data, claimants would need to survive for between two years (if the decision was on the last day of the year) and three years (if the decision was on the first day of the year).

implementation of SDM authority varies greatly across states, none of the DDS managers indicated that SDM authority affects the allocation of claims. In other words, where SDM authority exists, it is not used to define the availability of examiners in the electronic interface.

On balance, while nearly all claims were allocated randomly by computer to examiners on the basis of availability, a small percent were assigned nonrandomly. As a percentage of initial claims, however, the exceptions to the standard processes were a very small part of the caseload in 2005 and 2006. We estimate that all of the exceptions discussed here sum to less than 0.5 percent of our sample of applications. This low percent reflects the facts that most high-priority cases at this time were for Terminal Illness, and consequently, most of these individuals did not survive long enough to be included in our sample.