

The Work Disincentive Effects of the Disability Insurance Program in the 1990s¹

Susan Chen
Purdue University²

and

Wilbert van der Klaauw
Federal Reserve Bank of New York³

Forthcoming in Journal of Econometrics

Abstract

In this paper we evaluate the work disincentive effects of the Disability Insurance program during the 1990s using comparison group and regression-discontinuity methods. The latter approach exploits a particular feature of the DI eligibility determination process to estimate the program's impact on labor supply for an important subset of DI applicants. Using merged survey-administrative data, we find that during the 1990s the labor force participation rate of DI beneficiaries would have been at most 20 percentage points higher had none received benefits. In addition, we find even smaller labor supply responses for the subset of 'marginal' applicants whose disability determination is based on vocational factors.

JEL classification codes: I12, J22.

Keywords: regression-discontinuity design, disability, labor force participation.

¹The views and opinions offered in the paper do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System as a whole.

²Mailing address: Department of Agricultural Economics, Purdue University, 403 West State Street, Suite 591A, West Lafayette, IN 47907, Email: sechen@purdue.edu.

³Mailing address: Microeconomic and Regional Studies Function, Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045. Email: Wilbert.vanderklaauw@ny.frb.org.

1 Introduction

With labor force participation rates of older males falling throughout the last three decades, researchers have sought to explain this phenomenon by examining the interaction between a number of different social insurance programs and labor force participation (Leonard, 1986; Burtless, 1999; Bound and Burkhauser, 2000). Among these, the Disability Insurance (DI) program has been identified as being one of the primary potential reasons for the non-participation of prime aged males in the labor force. As its eligibility criteria imply a very high tax rate on earnings, the DI program has long been criticized for its apparent work disincentives.

Despite an extensive body of research on this issue, there is little consensus among economists on the magnitude of its work disincentives and on the role attributed to the DI program in explaining the large decline in the labor force participation rate of older men. A better understanding of the incentive effects of the disability insurance program is not only needed to explain DI's contribution to the changing employment rates of older men and women but also to improve our ability to predict, explain, and manage the increasing costs of DI programs, which is of great concern to decision makers, and will be essential for evaluating potential changes to the disability program.

Participation in the DI program is the outcome of an individual's decision to apply for disability benefits combined with an eligibility determination decision. To the extent that incentives to apply vary across individuals and that eligibility criteria depend on various individual characteristics, disability benefit receipt cannot be treated as an exogenous explanatory variable in a labor force participation equation. With generous income replacement ratios, particularly for low earners, there is an economic incentive for the disabled previously capable of work to stop working and for people who are not truly disabled but receive a high disutility from working to take advantage of the program (for example, by misreporting their health status). Similarly, medical and vocational criteria used to determine eligibility for disability benefits result in large differences in characteristics between those receiving and not-receiving DI benefits. Because some of these characteristics are likely to be unobserved by the econometrician, this implies that DI receipt should be treated as endogenous in an econometric analysis of its effect on labor supply.

A popular way to deal with this problem has been to model labor force participation (LFP) (or non-participation) as a function of the ratio of potential benefit levels to wages, known as the replacement rate. Most of the earlier empirical studies conducted in this area analyzed the impact of the DI program this way using cross-sectional data and traditional econometric regression models. In these models, non-participation is modeled as a function of the replacement rate and demographic and health characteristics such as age, education, and health status. In the best-known work of this type, Parsons (1980) estimated a non-LFP

elasticity for prime aged men (45 to 59) of 0.63, while Slade (1984) found an elasticity of 0.81.

Two problems arise in such an analysis. First, by grouping both wages and benefit levels into the replacement ratio, the separate impacts of wages versus benefit levels on non-labor force participation are confounded. Second, the actual benefit amounts participants receive depend on past earnings, and therefore on past work decisions, generating an additional direct source of endogeneity in the amount of benefits received.

In an attempt to address this endogeneity problem Haveman and Wolfe (1984a, 1984b) replaced the actual replacement rate with a predicted value obtained from a first stage regression of the replacement rate on a set of exogenous variables. In contrast to the earlier studies, they found much lower elasticity estimates of between 0 and 0.03. To identify the replacement rate effect (or the separate wage and disability benefit effects) some exogenous variables that determine wages or (and) disability benefits must be excluded from the labor force participation equation. However, without a convincing justification for their exclusion restrictions it is not clear how credible their estimates are.

While these earlier cross-sectional studies based on US data either ignored the potential endogeneity of the replacement rate or relied on arbitrary exclusion restrictions for identification, three recent studies explore alternative identification approaches for dealing with the endogeneity of disability benefit receipt. Gruber (2000) exploits an exogenous policy change conducted in Canada in 1987 where the benefit levels of the rest of the country were adjusted upwards to meet those of Quebec province. Using data covering the 1985-1989 period, he estimates the elasticity of labor force non-participation with respect to DI benefit levels to be between 0.28 and 0.36. The identification approach and the credibility of his estimate depend on the validity of the assumption that any changes in the relative labor market conditions in Quebec as compared to the rest of the country during this period were uncorrelated with the differential change in DI benefits.

Autor and Duggan (2003) also use differential time variation in average benefits across geographical regions to identify the impact of disability insurance on the labor force participation of low skilled workers. Using state level data from the CPS and the Social Security Administration, they exploit variation in the replacement rate due to differences across states and over time in the wage distribution, to identify the effect for low-income workers. They maintain that the widening dispersion of earnings in the US, combined with the progressivity of the disability benefits formula and the fact that DI benefits are set nationally and do not adjust for variation in regional wage levels, provide an exogenous measure of program generosity independent of a workers underlying taste for work. They conclude that the disability system provided many low-skilled workers with a viable alternative to unemployment. They estimate that the overall unemployment rate in 1998 would have been one half a percentage point higher in the absence of the DI program. Unfortunately, their reported estimates do

not allow calculation of an elasticity that can be compared to those in other studies. The identification strategy relies on the absence of other differences across states in both the changes in labor market conditions over time as well as the impact of such changes on labor supply, which seems problematic since variation in the wage distribution over time across states can itself be expected to directly affect labor supply.

A very different approach to deal with the non-comparability between DI recipients and non-recipients was suggested by Bound (1989). Instead of estimating the effect of benefits and wages on labor supply, Bound considers the more basic problem of evaluating the effect on labor supply of being a DI beneficiary. Arguing that they should be much more similar in observed and unobserved characteristics, he uses a sample of rejected disability applicants as a control group for DI beneficiaries and considers their labor force participation rate as an estimate of the counterfactual LFP rate of DI beneficiaries. Given that accepted and rejected applicants are not completely comparable, with rejected applicants generally being healthier, Bound argues that their work behavior forms an upper bound for the behavior of DI beneficiaries had they not been receiving benefits. Bound estimates in this way that the labor force participation rate of DI beneficiaries in his sample would have been at most 30 percentage points higher had they not received disability benefits.

The validity of Bound's identification approach relies on several assumptions. First, the interpretation of the LFP rate of rejected applicants as an upper bound is based on the assumption that the only difference between rejected DI applicants and beneficiaries is that the former are in better health. While most rejections are based on an initial medical screening, the disability determination process also has an important component, which will be discussed in detail later, which is based on vocational factors. As a result rejected applicants and beneficiaries may on average differ not only in their average health but also in other characteristics, such as their average pre-application earnings, work histories, age and education level. Moreover, they may differ in their preferences for working. It may be the difference in these average characteristics rather than DI benefit receipt that leads to the lower labor force attachment of beneficiaries.

Second, strictly speaking Bound's identification procedure provides an upper limit for the effect of disability benefit receipt on the labor force participation rate of applicants only. For it to represent an upper limit of the effect of the DI program on total labor supply, the behavior of rejected DI applicants should be comparable to what it would have been if the DI program had not existed. As pointed out by Parsons (1991), the LFP of rejected applicants may underestimate what their LFP would have been had they not applied, if denied applicants are awaiting appeals or are planning to reapply, or if the period of time they were out of work while applying for benefits makes it more difficult to return to work.

Bound (1989, 1991) directly addresses both issues and presents evidence suggesting that both are unlikely to invalidate his main findings. With respect to the second, he argues

that most appeals should have been accounted for by restricting his sample to applicants who filed their applications at least 18 months before the survey dates. He notes that even if a denied applicant were to exhaust all appeals, the maximum time between application and awards (including reconsiderations, appeals, etc.) would have been about 496 days in 1982. He does agree that some of the rejected applicants in his sample may be out of the labor force while planning to reapply. However, he argues that they only represent a small number of applicants who probably have a lower opportunity cost of remaining out of the labor force (those in worse health and unemployed). Any resulting bias is therefore likely to be small. Bound also argues that there exists little evidence that rejected applicants face special problems in returning to work due to the time they were out of work while applying for benefits.

While Bound's estimates are arguably among the most convincing to date, they were based on data collected in 1972 and 1978 and the DI program as well as the population of elderly has changed considerably since then. Over the last thirty years, the screening process for DI applicants has become more formalized, especially with respect to borderline cases. In addition, administrative control of DI has undergone many changes. Between 1985 and 2004 the number of disabled individuals receiving DI increased by over 100 percent. This increase is particularly striking given the enactment of the Americans with Disabilities Act in 1990. This act should have provided more economic opportunity for the disabled and relieved some of the pressure on growing DI rolls. In light of these macro economic changes, as well as in society's attitude towards people on welfare over the last 30 years, an analysis of the impact of DI program using more recent data is timely and important.

Using matched survey-administrative data on DI applicants from the 1990s, our estimation approach will consist of two separate components. We first apply Bound's comparison group approach to estimate an upper bound of the impact of disability benefit on labor force participation. We will discuss its underlying assumptions and their applicability to our data set. Second, we adopt a quasi-experimental approach to obtain a point estimate of the impact of disability benefit receipt for an important subgroup of applicants most likely to be the target of any policy reform: 'marginal' applicants who are not immediately awarded or denied benefits on the basis on their specific medical impairment and whose eligibility needs to be determined by considering vocational factors. The Regression-Discontinuity (RD) approach that we use to obtain this estimate, exploits the fact that the eligibility determination process is based in part on an individual's age. This provides us with an intuitive way of obtaining a policy-relevant estimate of the program's impact on labor supply.

While the literature to date has focused mainly on the labor supply of men, we extend our analysis to study the employment effects of the DI program on the labor supply of both men and women. In addition, we analyze both the short and long-term employment effects of benefit receipt.

Our estimates indicate that the work disincentive effects associated with DI benefit receipt during the 1990s were relatively modest, implying that the labor force participation rate of DI beneficiaries would have been at most 20 percentage points higher had they not received benefits. This result is robust to a variety of specification tests, appears insensitive to how much time has passed between the award date and the measurement of labor supply, applies to male and female applicants, as well as applicants to the SSDI and the SSI programs. We find even smaller effect estimates for a group of ‘marginal’ applicants whose disability determination was based on vocational considerations. However, within this group we find estimates to vary, with male and SSDI applicants generally showing somewhat larger labor supply responses than female and SSI applicants.

The rest of the paper is organized as follows. We begin with an overview of the disability determination process and the program’s benefit rules in section 2. This is followed by a description of our data set in section 3. In section 4 we apply Bound’s comparison group approach to estimate an upper limit on the impact of disability benefit receipt on the labor supply of DI applicants. We compare our results to those reported by Bound and discuss the validity of the conditions under which this estimate represents an upper bound. Section 5 describes a relatively unexplored feature of the DI program’s disability determination procedure, the medical vocational grid, and shows how the use of this grid leads to discontinuities in the award rate as a function of the applicant’s age. We explain how this can be used to identify and estimate the program’s work disincentive effect for the policy relevant subgroup of applicants whose disability status is determined by the grid. In section 6 we assess the sensitivity of our findings to the estimation method used, examine the validity of the underlying assumptions of our evaluation approach, and present estimates for several subsamples of DI applicants. We conclude in section 7 with a discussion of the implications of our findings.

2 The Disability Program

Cash assistance is provided to the disabled under two federal programs: the Social Security Disability Insurance (SSDI) Program and the Supplemental Security Income (SSI) program. The SSDI program is part of the Old-Age Survivors and Disability Program and is funded mainly through payroll taxes. It is a social insurance program with eligibility conditioned on previous sufficient employment in jobs covered by Social Security.¹ It was designed to provide partial earnings replacement to all workers under age 65 and its benefit amounts

¹To qualify for SSDI benefits, an individual must have sufficient work experience in employment subject to Social Security contributions. The amount of work experience depends on the age at which the person becomes disabled. Generally, a person needs to have worked 10 years, 5 of which during the 10 years preceding the year of disability. Younger workers may qualify with less work experience.

are dependent on previous earnings. In addition to paying benefits to disabled workers, the program provides supplemental benefits to dependent children of disabled workers, and in some cases also to the spouse.

In 1999, approximately 4.9 million disabled workers received SSDI benefits, of which 2.8 million were men and 2.1 million were women.² In addition, more than 1.5 million individuals received SSDI benefits as dependent family members of disabled workers, of which over 90% were children. Benefits are based on past earnings using a formula that provides lower earners with higher replacement rates. The average monthly benefit received by a disabled male and female worker in 1999 was respectively \$845 and \$740, and the average total amount received by a disabled person with eligible children was \$1273. At age 65 the benefit is switched to a retirement benefit but the amount of the payment remains the same. All workers contribute to the SSDI trust fund through their Social Security taxes.

Since 1972 cash assistance to the disabled has also been provided through the SSI program. In 1999, approximately 4.0 million individuals aged 18-64 received disability benefits under SSI, of which 55 percent were women. SSI provides a minimum level of income to needy blind, aged and disabled individuals and, unlike SSDI is not tied to previous employment. Instead it is subject to an earnings and assets test. In June 1999 this was earnings of less than \$500 per month (\$751 for jointly eligible couples) and assets not exceeding \$2000 (\$3000 for couples). SSDI recipients who meet the assets and income criteria may also receive SSI benefits, and in 1999 a little over a quarter did so. In June 1999 the standard monthly SSI benefit to an individual was \$500. Spouse and child benefits are not awarded to participants in the SSI program. If a child or spouse is disabled then they would apply for benefits under their own social security number. More than one member of a household can apply for benefits. In 1999 the standard benefit amount was capped if both husband and wife apply at \$751.

A person seeking benefits under either program applies for them at an office of the Social Security Administration (SSA). After the federal officials and the applicant have gathered sufficient information to complete the application, it is forwarded to a state disability determination service (DDS). Disability examiners in this office, with the aid of vocational and medical consultants, then determine whether a person is disabled based on whether they satisfy eligibility rules and meet the criteria laid out in the Code of Federal Regulations. The Code requires that each applicant go through a sequential Disability Determination process where medical factors as well as vocational factors (such as age, education and past employment) are used to determine an applicant's ability to work. The determination screening process consists of five stages. The sequential process ensures that easy cases are

²Unless otherwise indicated, all statistics in this section are taken from the Social Security Agency Annual Statistical Bulletin, 2000.

resolved in the preliminary stages, and benefits immediately issued, while the more difficult or marginal applicants proceed to later stages for determination. The process is summarized in the flowchart in Figure 1 which is taken from Lahiri et al., (1995).

An applicant to either the SSI or the SSDI program must go through this determination process in order to be awarded benefits. In the first stage, there is an earnings screen (earnings of more than \$500 in 1999). An applicant cannot be engaged in an activity, which is both substantial and gainful. Applicants who do not meet the earnings criteria are rejected. A very small number of people are rejected at this stage since the rule is very clear and can be found in most material distributed to the public by the Social Security Administration.

In the second stage, the severity of the medical impairment is assessed. An impairment is considered to be severe if it meets the duration test i.e. the impairment is expected to last at least 12 months, or result in death, and significantly limits the physical or mental ability of an individual to perform his work-related activities. Using numbers from Lahiri et al., approximately 18% of applicants whose applications were processed during the 1986-1993 period were rejected at this stage.

In the third stage, the medical evidence obtained on application is assessed using specific codified clinical criteria relating to both the nature and severity of the impairment. Applicants that meet these criteria are then deemed disabled on the basis of a Medical Allowance and accepted into the program. Applicants who do not meet these criteria then pass on to stage four. Lahiri et al. report that 36% of applicants who reach this stage are accepted on the basis of a medical allowance.

At stage four, the applicant's past work experience is used to determine whether he can perform in the job he had before the onset of the disability. The applicant's residual functional capacity is also determined at this time. The residual functional capacity is the ability that the applicant, though severely impaired, has to perform specific work related activities such as walking, lifting objects or taking instruction. If he is able to carry out his past job then he is denied; if he cannot perform his past job, then he moves to the next stage of the determination process. Lahiri et al. report that about 29% of applications that reach this stage are denied.

Stage five is the final stage in the disability determination process. At this point, applicants are either accepted or rejected. The determination now depends on whether the applicant can do any type of work in the economy. The residual functional capacity determined in the previous stage is used together with vocational factors (age, education, and work experience) to judge whether an applicant could be considered for alternative types of work in jobs other than those he has held. A grid is provided at this time to guide the disability assessor on how to make judgments. Lahiri et al. report that 37% of the original applicants to the SSDI program face the grid. The use of the grid will provide the basis of our regression discontinuity analysis and will be discussed in detail below.

Applicants who are initially rejected can request a reconsideration. Of the 57 and 69 percent of initial applications that were rejected in 1992 and 1995, respectively 47 and 49 percent did so. The reconsideration is carried out by a separate team of evaluators and is not considered to be a new application. In 1992 (1995) about 17% (13%) of the people who applied for reconsideration were granted benefits.³

If an applicant is rejected at the reconsideration stage, he can request a hearing before an Administrative Law Judge (ALJ). In 1992 (1995) about 63% (66%) of those denied benefits at the reconsideration stage advanced to this level and 69% (62%) were granted benefits. Further levels of appeals are conducted at the Social Security Appeals Council and in the Federal Courts. Of those denied benefits at the ALJ stage in 1992 and 1995 about 80% and 45% respectively requested a hearing before the appeals council. The award rate at this level for 1992 (1995) was 4% (3%). The final level of appeal is in the federal court. At this stage, 13% (18%) of those rejected requested a hearing before a Federal judge in 1992 (1995). Only 20% in 1992 and 10% in 1995 of those who appealed were granted benefits.

In summary about 14% (19%) of the initial applicant pool carried their claims beyond the reconsideration level in 1992 (1995), representing 27% (29%) of applications not awarded benefits by that stage. An important distinction between subsequent appeals and determinations made by the state disability determination service is that disability evaluators at these subsequent appeals are not bound by the five-stage process adhered to at the lower levels, and therefore have considerably more discretion in making award decisions. We will discuss this in more detail below.

Since the definition of disabled mandates a permanently disabling condition, very few DI beneficiaries ever leave the rolls. In both the SSDI and SSI programs beneficiaries may have their application reviewed and their benefits terminated if they have recovered enough and are no longer considered disabled. Since termination of benefits is such a politically sensitive instrument, the actual number of cases reviewed for termination has oscillated with the political climate. During the Reagan years 13.5% of cases were reviewed with as many as 40% of these beneficiaries losing benefits. By 1995, the number of reviews had dropped to only 1% with about 15% of the reviewed cases losing benefits (Rupp and Stapleton, 1998).

3 Data

In our analysis we use data from a newly constructed data set that merges the 1990-1996 panels of the Survey of Income and Program Participation (SIPP) with restricted Social Security data containing detailed SSDI and SSI application and award information. The two

³The sources for these 1992 and 1995 statistics and those reported below are the 1993 and 1996 Green Books.

datasets were exact matched for SIPP sample members who applied for disability benefits and whose applications were adjudicated between 1989 and 2000.

The restricted data on awards and applications comes from the Social Security Administration’s “831 file”. This file, as made available to us, contains detailed information on each transaction that is made at each stage of the disability determination process *excluding* adjudications made beyond the reconsideration stage. Of particular interest to our analysis is the data on both rejected and accepted workers who apply for disability insurance under the SSI or SSDI programs. The 831 file contains information on the award decision for both SSI and SSDI applicants, the date an application was filed, the date an application decision was made, the stage in the application process where the determination was made and whether the individual was granted or denied DI on the basis of vocational factors, and the type of claim applied for (child’s benefit, widow’s benefit etc.). In our analysis each applicant will be characterized by the date of the award decision. We are interested in the labor supply responses following the completion of a person’s application process (which may include requests for reconsiderations and re-applications), so we define this date to be the date at which the final determination up to the reconsideration stage was made.⁴

While an individual’s merged administrative data record may contain transactions data that precede and succeed the SIPP observation period, it is useful to distinguish between three cases that can occur when the SIPP and 831 file are merged:

(1) the application in the 831 file occurred during an individual’s SIPP panel observation period – the SIPP data is current with the DI application

(2) the application occurs before the person shows up in the SIPP file – the DI application predates the SIPP data

(3) the DI application occurs after the SIPP observation period available for this person has ended – the SIPP data is older than the DI application.

Our analysis of post-application employment outcomes is based on two different data sets both of which exclude the category (3) cases. That is, they include only matched SIPP-831 data on SIPP members for whom post-award decision employment information is available. The first of these data sets has a short post-decision time horizon and includes all individuals between the ages of 35 and 61 at the time of the award decision for whom we have any pre-age 62 employment information between 1 to 24 months after the decision date. This sample is created to study post award work decisions within a short time horizon after the award decision. The second dataset has a much longer time span and includes all individuals aged 35 to 61 at the time of the award decision for whom we have any pre-age 62 labor force information at least one month after the award decision. The time horizon ranges from 1

⁴The 831 file contains a non-negligible number of individuals who applied multiple times during the 1989-2000 period. For these individuals, the award date is defined as the final award decision date recorded in the 831 file for the individual’s most recent application.

month to 11 years. This longer time horizon will allow us to examine how applicants adjust their labor force behavior over time.

Some of our analysis will be restricted to the group of individuals whose disability determination was made on the basis of vocational considerations (at stage 5 of the disability determination process). This is the group of applicants whose disability determination could not be made solely on medical grounds, but instead was based on vocational criteria. This group is of special interest as it includes mostly marginal cases, whose disability classification can be expected to be a potential target of future policy reforms. In some of our analyses of benefit award decisions, we will also make use of case (3) data. In the latter case, the administrative data on all SIPP members who applied at any point during the 1989-2000 period are pooled to form, what we call, the “extended” sample.

The SIPP dataset provides information about the number of hours worked in each month of the survey. We examine the effect of DI on two employment measures: monthly hours of work and labor force participation. For the monthly hours of work variable we use the number of hours the individual worked in the last SIPP survey month available for this individual at which he or she was less or equal to 61 years old. Similarly, we define an individual to be a participant in the labor force if the number of hours worked in the last pre-age 62 SIPP survey month is greater than zero.

For reasons that will be described later, an important explanatory variable in our analysis is the individual’s age at the date of the award decision (measured in months) which we defined earlier. Other explanatory variables included in some of our analyses are indicators for whether the person’s highest educational attainment was a high school diploma, and whether the person obtained any college education. More specifically, an individual was defined to have a high school education with 12 years of education, and more than a high school education if the individual completed more than 12 years of education.

The descriptive statistics for key variables of interest are listed in Table 1, for the short-run and long-run samples, and also separately for the samples of applicants whose award decision was made on the basis of vocational factors (we call these stage 5 applicants in our figures and tables). The average age of the short run sample is a little under 50 years and about 62% of the sample has at least a high school diploma. A slight majority of applicants are female, and about 31% are nonwhite. Measured on average 18 months after the award date, about 14 percent of the short run sample work and among those who work, the mean number of hours worked is 33 hours per week. The long run sample is on average approximately half a year younger, is slightly less likely to have completed high school, and has a marginally higher labor force participation rate and average hours of work measured on average 3 years after the decision date.

When we only consider the subsamples of applicants whose award decisions were made at Stage 5 of the disability determination process, we find them to be surprisingly similar, the

main differences being that a somewhat smaller proportion of these applicants are female or attended college, and the labor force participation rates for this group in both the short and long run samples are slightly lower. The similarity in employment rates suggest that while cases determined in stages 1 to 4 are more likely to include those with very bad and relatively good health, on average their employment rates and demographic characteristics are similar to cases which reach stage 5. While the former group includes comparably sized groups of cases accepted and rejected on the basis of medical conditions, the stage 5 cases represent the more difficult to decide borderline cases. In our sample, the latter group constitutes 39% of all applicants.⁵

4 Evaluation with Rejected Applicants as Comparison Group

Before discussing estimates based on Bound’s comparison group approach, it is important to highlight some important differences between our dataset and that used by Bound. First, instead of relying on self-reported DI application and receipt data, we use administrative DI records. It has been argued that rejected applicants may be less likely to self-report a rejected application to the SSDI program if they subsequently joined the labor force. By using administrative data we avoid any potential biases that may arise due to such under-reporting of applications by rejected applicants.

Second, while Bound’s original study was based only on information about SSDI applicants, our analysis also includes applicants to the SSI program. In fact, 42% of the applicants in our long-run sample applied only for SSDI benefits, 28% applied only for SSI benefits and the remaining 30% applied for both. While we consider both programs together for efficiency reasons, it is important to recognize that there are some clear differences between the two groups of applicants that may affect the results. Bound et al., (2003) found that SSDI and SSI applicants are very different in terms of pre-application labor force participation and household income. They found that SSI applicants had lower average employment and labor earnings than SSDI applicants. In addition, they reported that rejected SSI applicants were less likely to work than rejected SSDI applicants.⁶ Finally, given these socio-economic differences between SSI and SSDI applicants, SSDI applicants are more likely to have stronger applications (reflected in stronger letters or evaluations from a doctor of their own choosing) than SSI applicants, who are on average less able to afford their own doctor and more likely

⁵A similar pattern was reported by Hu et al., (2001) who found that approximately 28% were awarded benefits and 33% were denied benefits during stages 1 through 4, with the remaining 38% reaching stage 5.

⁶Bound et al., (2003) use a dataset similar to ours but for different years; they matched the 1990-1993 SIPP panels to the 831 file administered between 1977 to 1997.

to be evaluated by a local Social Security Administration office.

Third, instead of restricting our sample to male applicants, our data includes both male and female applicants. The effect of disability benefit receipt on labor supply may differ between both groups, and also between SSDI and SSI applicants. To investigate this, and to compare our results more directly to those in Bound (1989), our analysis will include a separate analysis for male SSDI applicants.

There are also some less obvious differences between the two data sets in the measurement of labor supply and disability benefit receipt. Bound's sample included applicants who *applied* at least 18 months prior to the 1972 and 1978 survey dates at which their labor supply was measured, with median years of application for both samples being respectively 1969 and 1975. Our sample includes applicants for whom we observe employment data at least 1 month and up to 11 years after their *award decision* was made. In case a person was initially denied benefits and decided to request a reconsideration, this date would equal the date at which the reconsideration decision was made. Moreover, for those who applied multiple times during the 1989-2000 period, this date refers to the award date of the most recent application. In our 'extended sample' containing all SIPP members who applied for DI benefits during the 1989-2000 period, for 66% of the cases the award date corresponded to the initial application phase, while for 34% of cases it corresponded to the award decision at the reconsideration stage.⁷ For both cases combined we find that there was an average delay of 4 to 5 months between the date of application and the award decision date. However, it is important to note that over 18% of the first group of applicants in our sample had applied for DI benefits before, while some 22% of the second group had done so. When measured from the filing date of the first application, we find that on average 10 months had passed until our measure of the final decision date. In addition, as shown in Table 1, the average time between the decision date and the date at which the individual's employment status was measured was approximately 1.5 years for the short run sample and 3 years for the long run sample. Finally, it is important to recall that these records exclude adjudications made above the reconsideration stage. We discuss the implications of this aspect of our data later in the paper.

To better understand the two empirical approaches that we use in this study we begin with a discussion of the general evaluation problem. Let y_i be the outcome measure and t_i the treatment indicator, where $t_i = 1$ if treatment was received and $t_i = 0$ if not. Further, let $y_i(1)$ be the outcome if the individual is given treatment and $y_i(0)$ if not treated. We never observe a person simultaneously participating and not participating in the DI program.

⁷The award rates at the initial and reconsideration stages in our sample were, respectively 42% and 17% which are comparable to the rates reported for 1992 and 1995 in section 2. Of those denied benefits at the initial determination stage 58% requested a reconsideration, which is a little higher than the rates for 1992 and 1995.

What we observe is

$$y_i = t_i y_i(1) + (1 - t_i) y_i(0). \quad (1)$$

In our case, we want to evaluate the effect of disability benefit receipt t on labor force participation y . The evaluation strategy for this evaluation problem that was proposed in Bound (1989), is to use rejected applicants as a comparison group. To understand this approach, consider the evaluation of the average treatment effect on the treated $E[y_i(1) - y_i(0)|t_i = 1]$, or in our case the average employment effect associated with benefit receipt for DI beneficiaries. Estimation of this effect is hampered by the fact that while we can use the LFP of beneficiaries and non-beneficiaries in our sample to estimate $E[y_i(1)|t_i = 1]$ and $E[y_i(0)|t_i = 0]$, we do not observe the percentage of DI beneficiaries who would have worked in the absence of the program to estimate $E[y_i(0)|t_i = 1]$. John Bound argued that as rejected applicants are generally healthier than accepted beneficiaries, one could treat the observed employment rate of rejected DI applicants as an estimated upper bound on this missing counterfactual. That is, by restricting our sample to applicants, we are able to obtain an upper bound (in an absolute value sense) on the average treatment effect on the treated, because if $E[y_i(0)|t_i = 1] \leq E[y_i(0)|t_i = 0]$ then

$$\left| E[y_i(1) - y_i(0)|t_i = 1] \right| \leq \left| E[y_i(1)|t_i = 1] - E[y_i(0)|t_i = 0] \right|. \quad (2)$$

Note also that if, in addition, rejected applicants would have worked more as beneficiaries than current beneficiaries, i.e. $E[y_i(1)|t_i = 0] \geq E[y_i(1)|t_i = 1]$, then the right-hand side of (2) will represent an upper bound on the average treatment effect $E[y_i(1) - y_i(0)]$.

The statistics presented in Table 2 indicate that in both our long and short run samples, not more than approximately 22% of rejected applicants work. Those who work, work on average 34 hours per week. Of beneficiaries a little over 7% work in the survey month, and those who do, work on average 32 hours per week. Together these rates imply an upper bound on the reduction in the employment rate of applicants associated with disability benefit receipt of 14 and 15 percentage points for the short and long run samples, respectively. The corresponding standard errors of these estimates are 1.2 and 1.0 percentage points. When we restrict the samples to applicants whose disability status is determined by vocational rules, the estimated upper bounds (standard errors) are 15 (1.8) and 16 (1.6) percentage points, respectively. In terms of *monthly* hours of work, the estimated average reductions in labor supply (and standard errors) associated with benefit receipt are 21 (1.8) and 23 (1.7) hours for the short and long run samples of all applicants, and 22 (2.8) and 25 (2.6) for the short and long run vocational samples.

Comparing our estimate of the decline in labor force participation associated with disability benefit receipt to those in Bound (1989), we find our estimate to be somewhat smaller. Bound found that up to 33% of rejected applicants were working at the time of the survey,

and his estimates imply that benefit receipt led to a 29.4 percentage point reduction in the employment rate of applicants. There are several potential explanations for the difference in results.

First, while Bound's data were collected in the 1970s, our data were collected in the 1990s. Over the two decades between Bound's analysis and ours the DI program has been liberalized affecting the eligibility of many applicants.⁸ Decreased opportunities for low-skilled workers in the labor force during the early and middle nineties coupled with a decreasing real wage for this segment of the labor force, a relative increase in benefits, as well as a higher acceptance rate for DI benefits are likely to have reduced the cost of applying and induced such workers to apply for disability benefits (Juhn et al., 2002). Autor and Duggan in fact argued that this expansion of the DI program actually helped lower the unemployment rate in the US by 0.5 percentage points, and by 4.2 points for high school dropouts. The resulting compositional change in the pool of applicants to the DI program, with fewer opportunities, lower labor market attachments and willingness to work, could help explain our finding of a smaller negative employment effect of benefit receipt for DI applicants during the nineties.

Second, as discussed earlier there are several differences between the samples which may account for the difference in estimates. Bound's sample only included male applicants aged 45-64, while our sample includes both male and female applicants aged 35 to 61. Our sample also includes individuals who applied only to the SSI program, a group that constitutes 28% of our total sample. As this group of individuals includes those ineligible for SSDI benefits due to insufficient work experience in covered employment, one may expect a lower labor force attachment for this group. We will analyze the role of these differences in sample composition in section 6.

Third, a potentially important concern for our analysis is that because our administrative data lack information about appeals beyond the reconsideration stage, some beneficiaries in our data may be mistakenly classified as rejected applicants, which in turn could lead to a downward bias in our estimates. To assess the magnitude of this bias we need to estimate the fraction of rejected applicants in our data who may in fact have gone on to make successful appeals. Among all applicants to the SSDI and SSI programs who were denied benefits at the reconsideration stage in 1992 (1995), 63% (66%) filed further appeals, of whom 71% (63%) eventually were successful (Green Book, 1993, 1996). These national award and appeal rates imply that approximately 45 percent of all applicants whose applications were rejected at the reconsideration rate ended up receiving benefits. For our extended sample of 1990-1996 SIPP cohorts sample we do not have specific information about appeals beyond the

⁸For example, the number of beneficiaries with lower mortality diseases such as mental illness and back problems which are harder to diagnose and to verify has increased substantially. Between 1981 and 2000 the percentage of beneficiaries with mental and musculoskeletal disorders almost doubled from 27% in 1981 to over 50% in 2000.

reconsideration stage, but know that 42 percent of those rejected at the initial application stage did not ask for a reconsideration (and therefore did not file any further appeals), which in turn represents 47 percent of all rejected applicants in our sample.⁹ Therefore, based on an ultimate award rate of 45 percent for those rejected at the reconsideration stage, we predict that 24% (0.45×0.53) of the rejected applicants in our sample (which includes both those who did and did not ask for a reconsideration) to have become beneficiaries through subsequent appeals.¹⁰

If 24% of rejected applicants in our sample are in fact misclassified beneficiaries, then with true (rather than observed) benefit receipt denoted by the indicator t^* our estimate Δ of the right hand side of (2) is an estimate of $E[y|t = 1] - E[y|t = 0] = E[y|t^* = 1] - 0.76E[y|t^* = 0] - 0.24E[y|t^* = 1]$, which implies an estimated upper bound $E[y|t^* = 1] - E[y|t^* = 0]$ equal to $(1.0/0.76) * \Delta = 1.32 * \Delta$.¹¹ Thus we estimate that this misclassification leads to a downward bias in our estimates, with corrected estimates being approximately 32% larger than those reported earlier. As reported in the two bottom rows of Table 2 this would lead to estimated reductions in the labor force participation rate of between 19 and 21 percentage points and a reduction in average monthly hours of work of 28 to 33 for the various subsamples considered, estimates which are closer but still smaller (in absolute value) than Bound's estimates.

In addition to these point estimates, we can also compute an upper bound on the bias due to the misclassification of treatment status. As discussed earlier at most 53% of the rejected applicants in our sample could potentially have been misclassified beneficiaries. This implies an estimated upper bound (in absolute sense) that equals $(1.0/0.47) * \Delta = 2.15 * \Delta$ which corresponds to estimates of -30 and -32 percentage points respectively for the short and long run samples. These effect estimates are slightly larger than that of Bound, but are based on the extreme assumption that *all* applicants in our sample whose applications were rejected at the reconsideration stage, filed appeals of which *all* were successful.

The validity of Bound's comparison group approach rests on the assumption that the only difference between rejected DI applicants and beneficiaries is that the former are on average in better health. Given the nature of the disability determination process described

⁹As our administrative data on DI applications include all applications between 1989 and 2000, for a small proportion of cases in our sample, those with initial application dates close to the end of 2000, our award decision date and outcome may not include possible reconsiderations that may have been filed after 2000, implying that the misclassification rate may be slightly higher.

¹⁰Based on 1992 (1995) appeals and award rates, we know that approximately 42 (47)% of all unsuccessful applications up to the reconsideration stage had gone beyond the initial application stage, implying that 19 (20)% of the rejected applications in those cohorts were ultimately successful.

¹¹An implicit assumption made in these computations is that $E[y|t, t^*] = E[y|t^*]$ for all t, t^* , which requires award misclassification to be random, independent of potential outcomes $y(t^*)$. That is, it is assumed that the average labor supply of misclassified beneficiaries is the same as that of correctly identified beneficiaries.

earlier, it is obvious that rejected applicants will on average be in better health. However, they are also likely to differ in other characteristics. For example, the vocational criteria in Stage 5 (which will be discussed in greater detail in the next section) imply that those rejected at that stage will on average be younger and have higher education levels. This is indeed confirmed by Table 2 which shows rejected applicants on average to be younger and slightly more likely to have completed high school. As we would expect individuals with those characteristics to be more capable of working (all else equal), this difference in characteristics will therefore reinforce Bound's argument based on better health status, that the average labor supply of rejected applicants can be considered an upper bound for that of beneficiaries, had they not been awarded benefits.

While these differences in education and age may reinforce the validity of the estimation strategy, it is important to consider that beneficiaries and rejected applicants may differ in other characteristics. For example, as shown in Table 2 rejected applicants are significantly more likely to be nonwhite and female, and as a result this group may have lower, rather than higher potential employment rates than beneficiaries would have had without benefits. To analyze the implications of these differences in average observed characteristics for our estimates, we control for them using a regression analysis. The rows labelled 'conditional' in Table 2 report OLS estimates of the coefficient on DI benefit receipt, from labor supply regressions which include controls for age at time of the award decision, education, sex and race. The estimates are generally slightly smaller than those reported earlier, indicating that the small negative effects we attribute to disability benefit receipt are not due to differences in these observed characteristics.

Of course there may be differences in other characteristics that could make rejected applicants less, rather than more likely to work than beneficiaries in absence of benefit receipt. At a given age, race, gender and education level, those of above average health who decide to apply for benefits are more likely to have a lower cost or larger expected benefit from doing so. Bound (1989) notes that in general rejected applicants have slightly less work experience. In addition, it is not implausible that, all else equal, rejected applicants have a higher disutility from working than beneficiaries. The validity of Bound's approach therefore depends on the magnitude and importance of these differences. The evidence presented in Bound (1989) and Bound and Waidmann (1992) show very large differences in average health status between rejected applicants and beneficiaries which are likely to dominate the differences in work experience and unobserved tastes for work. There is however no way to determine the extent to which the two influences counteract each other. It is important to point out that while differences in unobservables may be a potential problem for this comparison group approach, this is not the case with the second evaluation strategy pursued below, the RD approach.

As mentioned earlier, another potential concern with Bound's analysis is that high non-

participation rates for rejected applicants may partly reflect the fact that some of these individuals were either still in the process of appealing or re-applying to the DI program. Because our analysis considers the employment outcomes of applicants some time after the reconsideration phase, and after they have re-applied, we should have few rejected applicants not working for these reasons. Moreover, as the time between the award and the measurement of employment status increases, the proportion of applicants who continue to be non-employed while continuing with appeals, should decline. The estimates for the long-term sample however are not very different from those for the short run sample, providing additional evidence that this is unlikely to be an important issue in our analysis.¹²

5 Evaluation Using A Regression Discontinuity Approach

As discussed in section 2, the disability determination for about 40% of applicants for whom it is difficult to determine eligibility on medical grounds, is based on vocational factors with the help of a grid. The grid was formalized and appeared as rule in 1978. Through a formula, the grid regulations relate certain worker characteristics such as age, education, and past work experience to the individual's residual functional capacity to perform work-related physical and mental activities. A simplified version of it is presented in Table A1 of the appendix. A more detailed outline can be found in the Code of Federal Regulations, Appendix 2 of Subpart P of the Medical - Vocational Guidelines. In the Guidelines, individuals are characterized into different age groups (under age 45, 45 to 49, 50 to 54, and 55 and over) as well as the residual functional capacity of the worker. The latter is defined using categories such as sedentary, light or medium, and is determined in Stage 4 of the disability determination process on the basis of medical conditions, and the person's experience and skill level.

As an example of how the grid is used, consider an applicant who has less than a high school education (7th through 11th grade), is (semi) skilled and who cannot easily enter into another profession (no entry into other work), and whose disability limits her to light work. Then according to the excerpt from the grid presented in Table 3 the applicant would be accepted if she was 55 or older at the time of the disability decision. However, if she was less than 55 years of age at that time, she would be rejected. Application of the grid in making award decisions would therefore lead to discontinuities in the award rate as a function of age. We have placed an asterisk beside all such age cutoffs in the grid, both in Table 3 and, for the full grid, Table A1.

¹²As the average award date in the long-run sample precedes that for the short-run sample, the comparison between short and long-run effects may be confounded by cohort effects. However, Bound (1989) similarly found that during the 1970s the labor force participation rates of rejected applicants did not rebound over time, suggesting that our findings are not due to cohort effects.

Conversations with staff at the Social Security Agency (SSA) and at the North Carolina Disability Office confirmed that the grid is indeed systematically used in the Stage 5 decision process. The SSA developed the vocational grid to reduce the subjectivity and lack of uniformity in applying vocational factors (2000 Green Book). While the grid is within the public domain, it is a little difficult to find and somewhat tedious to decipher for a person not familiar with the terminology used in the disability determination process. Except for Lahiri et al., (1995) who briefly mention the existence of the grid in their outline of the disability determination process, no economic research to date has discussed or used the grid in studying the disability program. Furthermore, in a thorough search of the information provided to the public on how to apply for disability benefits or advocacy information on how to secure benefits, the existence of a formal grid which uses age categories was never mentioned. Instead there are frequent references to general education-vocational guidelines without further details. The fact that applicants for disability benefits are unaware of the grid and can therefore not exploit or manipulate the fine details of the rule, considerably enhances the credibility of the RD approach we will adopt.

To see how discontinuities in the disability determination rule can be used to identify and estimate the work disincentive effect of disability benefit receipt, consider the hypothetical case where eligibility was solely determined on the basis of age where those below a given cutoff age, say 55 years, are ineligible and those at or above this age are eligible. Within a small age interval around 55 years, disability determination will be similar to a randomized experiment applied to individuals close to age 55. As the individuals are of similar age, those just below the cutoff age can be expected to be comparable to individuals just above the cutoff age in all characteristics. They also can be expected to have similar labor supply responses when receiving and not receiving disability benefits. In that case the only difference between the two groups will be that the applicants who are just below age 55 at the time their disability status is determined will not receive DI while those 55 or older will. As a result, the average labor supply of individuals just below the cutoff age in this example would represent a credible estimate of the missing counterfactual – what the labor supply of those just above the cutoff would have been had they not received DI benefits.

As outlined above, the disability determination made in Stage 5, in addition to being dependent on age also depends on other factors that we do not observe in the data such as residual functional capacity and illiteracy.¹³ Moreover, while the grid is regularly used in the determination process, there are cases where the recommendation made by the grid is overruled by the administrator. In other words, acceptance into the DI program also depends

¹³Note that the grid also contains discontinuities with respect to the highest grade completed. Unfortunately given the distribution of years of education in our samples, only a very small proportion of individuals in our samples had characteristics that would have placed them near these discontinuities. In addition, the information contained in the SIPP and 831 file on years of education appeared to be quite noisy.

on additional unobserved factors. As a result, instead of being a deterministic function of age, the award decision made at Stage 5 of the application process can be characterized by a probability function that will be discontinuous in age at the three cutoff values found in the grid at ages 45, 50 and 55. The assignment or selection process therefore conforms to that of the so-called fuzzy regression discontinuity design, where the award rate (the probability of receiving benefits) as a function of the age at application, is discontinuous at known age levels. The size of each discontinuity will depend on the fraction of individuals around each cutoff age with characteristics (such as education level and residual functional capacity) for which age can influence the disability determination in the grid, as well as the strictness in adhering to the grid in making award decisions. As will be shown below, similar to the example with a deterministic eligibility rule, we can estimate the effect of benefit receipt on labor supply by comparing the employment outcomes of applicants in a neighborhood just left and right of the age cutoff point.

The relationship between DI awards and age is explored in Figures 2 and 3. These graphs are estimated spline smooths of the award decision against the individual's age at the time the award determination was made. Both graphs are based on our 'extended' sample, which includes all 1990-1996 SIPP participants who applied sometime between 1989 and 2000. The vertical lines in each of these figures represent the Social Security grid cutoff ages of 45, 50 and 55. Figure 2 shows the spline smooth for all the applicants to the DI program. This plot includes both those whose eligibility decision was made on medical grounds in stages one to four, as well as those whose award decision was made at Stage 5 on the basis of vocational factors. While the graph reveals increases in the award rate around ages 50 and 55, they are relatively small. As the grid only comes into play for about 40% of the applicants in our sample, this is of course not surprising.

The rules imposed by the grid become much more apparent when the data is split into groups of applicants whose eligibility was determined on the basis of vocational versus medical criteria. As shown in Figure 3, for applicants whose status was determined in the first four stages of the determination process, we see no sharp increases in the award rate at the cutoff points. Clear jumps are seen, however, at the age cutoffs of 50 and 55 for the sample of applicants facing the grid, with the largest jump occurring at age 55. The fact that the sharp increases at these age cutoffs do not appear in the Stage 1 to 4 sample, provides credence to our claim that the jumps seen for the vocational sample are due to the use of the grid. They also indicate that the grid is applicable to a large proportion of applicants aged 55, a smaller proportion at age 50, and appears to apply to a negligible proportion of applicants in our data with ages near 45.

Even though the displayed award rates were smoothed (and therefore less likely to show sharp discontinuities), the main features of these spline smooths compare closely to those observed when we display raw award rates by yearly age intervals. As shown in Figure 4,

while the award rate for stage 5 applicants is generally increasing in age, the yearly award rate reveals clear discontinuities at ages 50 and 55.¹⁴ Also shown in the figure are predicted award rates from an estimated piecewise linear probability model specification of the award decision, estimates of which are presented in Table 4. The estimated jumps (and standard errors) in the award rate at the three cutoff ages were -0.016 (0.040) at age 45, 0.093 (0.045) at age 50 and 0.287 (0.037) at age 55. The table also presents estimates of the same specification for the short and long-run samples. They show similarly sized jumps in the award rate of 7 to 12 percentage points for the age 50 cutoff and 22 to 31 percentage points for the age 55 cutoff, with the latter always being statistically significant at the 99% level.

To see how discontinuities in the award rate can be exploited to estimate the work disincentive effect of benefit receipt, consider the general problem of estimating a causal effect of treatment t_i (DI benefit receipt) on outcome y_i (labor supply). Starting with the same notation from above, we can write:

$$y_i = \beta + \alpha_i t_i + u_i. \quad (3)$$

where $y_i(0) = \beta + u_i$ and $y_i(1) - y_i(0) = \alpha_i$.

Consider first the simple case of a sharp RD design (Thistlethwaite and Campbell, 1960), where treatment is a deterministic function of the individual's age A , as in $t_i = 1\{A_i \geq \bar{A}\}$. A person's age is likely to be directly correlated with labor supply y , thereby confounding the effect of t_i on y with that of age A . The idea of comparing average outcomes just to the right and the left of the cutoff value \bar{A} , can now be formalized with the help of two continuity assumptions:

Assumption S1: $E[u_i|A]$ is continuous at \bar{A} .

Assumption S2: $E[\alpha_i|A]$ is continuous at \bar{A} .

In the case of varying treatment effects, as long as $E[u_i|A]$ and $E[\alpha_i|A]$ are continuous in A at \bar{A} it follows that

$$\lim_{A \downarrow \bar{A}} E[y_i|A] - \lim_{A \uparrow \bar{A}} E[y_i|A] = E[\alpha_i|A = \bar{A}]. \quad (4)$$

The RD approach therefore identifies an average treatment effect $E[\alpha_i|A = \bar{A}]$ for individuals whose ages are close to \bar{A} . The continuity assumptions formalize the conditions discussed earlier which assume that individuals just above and below the cutoff age are comparable in characteristics, and have similar average labor supply responses when receiving and when not receiving disability benefits.

¹⁴For confidentiality reasons, given the sample sizes involved, we were not able to report award rates by narrower age intervals. Predicted award rates based on an estimated piecewise linear specification, as well as raw award rates by yearly age intervals for the short- and long-run samples are comparable to those shown in Figure 4. These figures can be found in Chen and van der Klaauw (2005), the working paper version of this article.

In our case the disability determination made in Stage 5, in addition to being dependent on age also depends on other factors that we do not observe in the data such as residual functional capacity and illiteracy. As a result, instead of being a deterministic function of age, the award decision can be characterized by a probability function that is discontinuous in age at the cutoff values found in the grid. As previously illustrated in Figure 4, the benefit award rate in our application of a fuzzy RD design has two clear discontinuities at ages 50 and 55. To see how it is still possible to identify an average treatment effect of interest in this case, consider first the case where the treatment effect is a constant ($\alpha_i = \alpha$). Then, at an age cutoff \bar{A}

$$\lim_{A \downarrow \bar{A}} E[y|A] - \lim_{A \uparrow \bar{A}} E[y|A] = \alpha \cdot \left[\lim_{A \downarrow \bar{A}} E[t|A] - \lim_{A \uparrow \bar{A}} E[t|A] \right] + \lim_{A \downarrow \bar{A}} E[u|A] - \lim_{A \uparrow \bar{A}} E[u|A]. \quad (5)$$

It follows that under assumptions S1 and S2 the common treatment effect α is identified by

$$\frac{\left(\lim_{A \downarrow \bar{A}} E[y|A] - \lim_{A \uparrow \bar{A}} E[y|A] \right)}{\left(\lim_{A \downarrow \bar{A}} E[t|A] - \lim_{A \uparrow \bar{A}} E[t|A] \right)}. \quad (6)$$

Note that the denominator is always nonzero because the propensity score has a discontinuity at the cutoff age \bar{A} .

As shown by Hahn et al., (2001), in case of varying treatment effects, under assumptions S1 and S2, and a local conditional independence assumption (requiring t_i to be independent of α_i conditional on A near \bar{A}), the ratio above identifies the average treatment effect for cases with values of A close to \bar{A} , $E[\alpha_i|A = \bar{A}]$. Furthermore, they show that this local independence assumption can be replaced by a weaker local monotonicity assumption, similar to that assumed by Imbens and Angrist (1994), which leads to the identification of a local average treatment effect (LATE):

$$\frac{\lim_{A \downarrow \bar{A}} E[y|A] - \lim_{A \uparrow \bar{A}} E[y|A]}{\lim_{A \downarrow \bar{A}} E[t|A] - \lim_{A \uparrow \bar{A}} E[t|A]} = \lim_{e \downarrow 0} E[\alpha_i | t_i(\bar{A} + e) - t_i(\bar{A} - e) = 1] \quad (7)$$

where $t_i(A)$ represents individual i th treatment assignment given any age A . This treatment effect represents the average treatment effect of the ‘compliers’, that is the subgroup of individuals for whom treatment changes discontinuously at the cutoff age. In our case this represents the population of individual with ages close to an age cutoff whose disability status as determined by the grid is dependent on whether their age is just below or above the cutoff.

To estimate the average treatment effect defined above, we adopt two different estimation methods: the two-stage method proposed by Van der Klaauw (2002) and the ‘local Wald’ estimator proposed by Hahn et al., (2001). In case of the latter, each of the limits in (6) are estimated using a one-sided uniform kernel estimator. More specifically, the limits are estimated as,

$$\begin{aligned} \lim_{A \downarrow \bar{A}_j} E[y_i|A] &= \frac{\sum_{i \in S} y_i w_i}{\sum_{i \in S} w_i}, & \lim_{A \uparrow \bar{A}_j} E[y_i|A] &= \frac{\sum_{i \in S} y_i (1-w_i)}{\sum_{i \in S} (1-w_i)}, \\ \lim_{A \downarrow \bar{A}_j} E[t_i|A] &= \frac{\sum_{i \in S} t_i w_i}{\sum_{i \in S} w_i}, & \text{and} & \quad \lim_{A \uparrow \bar{A}_j} E[t_i|A] &= \frac{\sum_{i \in S} t_i (1-w_i)}{\sum_{i \in S} (1-w_i)}, \end{aligned}$$

where w_i is an indicator for whether the observation is above or equal to the age cutoff such that $w_i = I\{\bar{A}_j \leq A_i < \bar{A}_j + h\}$ where h is a bandwidth, and S denotes the subsample around the cutoff point defined by $(\bar{A}_j - h \leq A_i < \bar{A}_j + h)$. This nonparametric estimator is therefore numerically equivalent to a Wald estimator applied to the discontinuity sample S , where t_i is instrumented by w_i .¹⁵ Instead of optimally choosing the bandwidth, we will provide estimates for three different interval widths.

The two-stage estimation procedure proposed by Van der Klaauw (2002) instead uses the entire sample of vocational applicants and involves the estimation of a control function augmented labor supply equation in which the DI program participation indicator is replaced by an estimated propensity score. More formally, in the first stage, the propensity score is estimated as,

$$E[t_i|A_i] = Pr(t_i = 1|A_i) = g(A_i) + \sum_{j=1}^3 \gamma_j \cdot 1\{A_i \geq \bar{A}_j\} \quad (8)$$

where $g(A_i)$ is a flexible continuous function in A_i , and the γ_j represent the discontinuities in the award rate due to the rules of the DI program. In the implementation of this approach we specify $g(A_i)$ to be a continuous piecewise linear function of A , as in $g(A) = \psi_{00} + \psi_{01}A + \sum_{j=1}^3 \psi_{1j}(A - \bar{A}_j)1\{A_i \geq \bar{A}_j\}$.

The estimated propensity score is used in the second stage to estimate the effect of DI on labor force participation. Following Heckman and Robb (1985), we include a control function specification $k(A_i)$ for the conditional mean $E[u_i|A_i]$ to control for the potential association between age (the selection variable) and employment status (outcome variable). The second stage equation is given by

$$y_i = \beta + \delta t_i + k(A_i) + v_i, \quad (9)$$

where in estimation t_i is replaced by the first stage estimate of $E[t|A_i]$. The continuous control function $k(A)$ is estimated semi-parametrically, using a power series approximation $k(A) \approx \sum_{j=1}^J \eta_j \cdot A^j$, where the number of power functions, J , is estimated from the data by Generalized Cross Validation (GCV) as in Newey et al., (1990).

Compared to the local Wald estimator discussed earlier, which only uses data in a relatively narrow interval around each cutoff age, the two-stage estimator uses more data while imposing additional smoothness assumptions on $E[u|A]$ and $E[\alpha_i|A]$. Not only does it impose continuity of these conditional expectation functions as function of A over their whole

¹⁵However, while numerically equivalent to a local Wald estimator, inference based on this estimator will be different from that based on a Wald estimator (see Hahn et al, 2001).

domains, it also assumes them to be differentiable over their domains.¹⁶ As discussed in Van der Klaauw (2002), in our case with three potential discontinuity points, the estimate of δ represents an estimate of a (weighted) average of the three local average treatment effects defined in (7) for the age 45, 50 and 55 cutoffs.

Table 5 presents regression discontinuity estimates of the impact of DI receipt on labor force participation and hours of work. The two-stage control function estimates reported in the first column of Table 5 (referred to as CFE) are based on a first stage piecewise linear specification of the award equation, estimates of which are reported in Table 4.¹⁷ The short-run sample estimate of a 20 percentage point decline in the LFP associated with DI benefit receipt is somewhat higher than the upper bound LFP estimate of 14 percentage points presented earlier in Table 2. However, the standard error is small enough to exclude, using a 95% confidence level, effects bigger (in absolute terms) than 34 percentage points.

In comparing these estimates it is important to recall, that they apply to two different subsamples of DI applicants. While the CFE is an estimate of a local average treatment effect (LATE) that applies to stage 5 applicants with ages near the discontinuity points (primarily 45 and 55), the comparison group estimate presented earlier in Table 2 is an estimated upper bound on the average treatment effect for *all* applicants. Another difference between the RD and comparison group estimates is that while the former approach accounts for selection on unobservables, the latter does not.

Somewhat surprisingly, we find the estimate of a 6 percentage point drop in LFP for the long-run sample to be smaller in absolute size than that for the short-run sample, and it also is smaller than the upper bound estimate of a 15 percentage point drop for this sample reported in Table 2. This suggests that allowing individuals more time for re-entry into the labor force is not very important, and is an indication that the proportion of individuals in our sample who may still be in the process of appealing previous denial decisions beyond the reconsideration stage (and who would return to work after final dismissal of their application) is relatively small. An alternative explanation, suggested by Bound (1991), would be that those individuals who continue their appeals are likely to be those with relatively low costs of applying, and with a low commitment to the labor force. The smaller longer term effect may also simply indicate that the disincentive effect associated with DI benefit receipt declines with the age at which labor supply is measured.

Also presented in Table 5 are local Wald estimates for the age 55 cutoff.¹⁸ Compared

¹⁶The Robinson-type estimator proposed by Porter (2003) similarly assumes continuity and differentiability of $E[u|A]$ and $E[\alpha_i|A]$.

¹⁷Estimates based on a piecewise quadratic specification yielded very similar results. The optimal order of the series for all control function estimates reported in the table was estimated to be 1, corresponding to a linear control function specification $k(A) = \eta_1 A$.

¹⁸We also compared individuals at the other two age cutoffs of 45 and 50. These results are not presented

to the estimates obtained with a comparison group approach, the local Wald estimates are slightly smaller and imply that benefit receipt is associated with a drop in employment of between 5 and 11 percentage points. While the estimated short-run LFP effect of DI receipt is somewhat smaller than that estimated using the two-stage control function approach, the local Wald estimates and CFEs based on the long-term sample are comparable.¹⁹

Similar results are found when comparing estimates of the effect of benefit receipt on monthly hours of work. As shown in the second panel of Table 5, the local Wald estimates, as well as the control function estimate for the long run sample, are generally slightly smaller than the comparison group estimates reported in Table 2, while the control function estimate for the short run sample is instead a little higher. The estimates indicate that without benefits applicants would on average have worked an extra 12 to 32 hours per month. Again we find the local Wald estimates for the short run sample to be smaller than the two-stage estimate, while both estimates provide similar estimates for the long run sample. Overall the estimates imply that, depending on the particular estimation method used, the work disincentive effects associated with DI benefit receipt for the marginal group of stage 5 applicants were either slightly lower or slightly higher than those obtained using Bound's comparison group approach, but with all estimates representing modest impacts.

Estimates of 'reduced form' specifications of the labor force participation equation, as well as the pattern of computed participation rates by yearly age intervals for both samples are consistent with this result. For example, Figure 5, shows the fitted 'reduced form' as well as the participation rates by age for the short-run sample. This figure, as well as that for the long-run sample (not shown), provides little evidence that the jumps in the award rates at the age cutoffs were accompanied by large changes in the labor force participation rates. This, together with corresponding parameter estimates and F-tests for the joint significance of the three jumps at ages 45, 50 and 55 for the reduced form participation equation, corroborates our finding of overall relatively weak labor supply effects of DI receipt during the 1990s.²⁰

6 Sensitivity Analysis

While the RD approach is often interpreted and implemented as an instrumental variable approach (as exemplified by the popularity of the local Wald estimator), more generally

because the numbers of observations on which they were based are too small to obtain precise enough estimates.

¹⁹It is important to keep in mind that the local Wald estimates and CFEs relate to slightly different subsamples, where the former represents a LATE for stage 5 applicants near age 55, while the latter applies to stage 5 applicants with ages close to 50 and 55.

²⁰Tables reporting estimates of the reduced form participation equation for both samples, as well as the long-run sample equivalent of Figure 5, can be found in Chen and van der Klaauw (2005).

nonparametric identification and estimation of treatment effects in the case of an RD design is based on the estimation of the four limits of conditional mean functions which comprise the numerator and denominator in (7). While alternative estimation approaches rely on different assumptions regarding these conditional mean functions, their common goal is to obtain consistent estimates of the discontinuities in the functions $E[y|A]$ and $E[t|A]$ at \bar{A} . While the nonparametric estimators discussed by Hahn et al., (2001) and Porter (2003) allow the functional forms of these two functions to differ from each other, as well as to be different on either side of the cutoff, more parametric approaches including the local Wald estimator as implemented here and instrumental variable approaches such as those used by Angrist and Lavy (1999) usually assume the functional forms to be the same. The reliability of estimates based on these latter approaches is tied to the validity of the parametric restrictions.

It is therefore important to analyze the sensitivity of parametric RD estimates to alternative specifications. For the local Wald estimator, which assumes average labor supply as well as the award rate to be unrelated to the applicant's age within a small age interval around the cutoff, this was done by varying the size of the interval width. As we would expect average labor supply to decline with age and the award rate to increase with age, we would expect increases in the interval width to be associated with a larger downward bias in the effect estimates. The estimates reported in Table 5 show that while the estimates are generally increasing (in absolute value) with interval width, they are reasonably stable. In fact, using formal Hausman tests based on pairwise comparisons of the three estimates, we could not reject the hypothesis that they are the same.

In case of the two-stage control function approach, which allows different specifications for $k(A)$ in the outcome and $g(A)$ in the first-stage benefit award equations, we reduce the risk of misspecification by adopting a series approximation of $k(A)$. As mentioned earlier, control function estimates based on a piecewise quadratic specification of the award equation were found to be almost identical to those for a first-stage piecewise linear specification. In analyzing the sensitivity of the two-stage estimates to the specification of the control function in the outcome equation we first compared estimates based on linear and quadratic specifications of $k(A)$, where the former represents the optimal specification as determined by cross-validation. We found them generally to be comparable, but with long-run estimate for the quadratic specification being larger and closer to the short-run estimate for both specifications. We also considered a continuous piecewise linear specification of the control function, where $k(A) = \eta_{00} + \eta_{01}A + \sum_{j=1}^3 \eta_{1j}(A - \bar{A}_j)1\{A \geq \bar{A}_j\}$. The two-stage estimation approach in this case is equivalent to two-stage least squares, where the jumps in the award rate at the three cutoff ages (the indicators $1\{A \geq \bar{A}_j\}$) serve as instruments. We found the employment effects associated with benefit receipt to be less precisely estimated in this case, but based on a 95 percent confidence interval, we can exclude declines in the labor force participation rate larger than approximately 28 and 40 percentage points for the short and

long-run samples respectively.²¹

Finally, we also pursued an estimation approach comparable in nature to that adopted by Angrist and Lavy (1999). In our implementation of this approach, we first use the larger ‘extended’ vocational sample to estimate a piecewise linear specification of the award equation, estimates of which were reported in Table 4. These estimates were then used to compute predicted award rates for the applicants in our short and long-run vocational samples. This was followed by TSLS estimation of the control-function augmented labor supply equations relating employment and hours of work to benefit receipt and a linear (quadratic) specification in age, where benefit receipt is instrumented using the predicted award rate and the linear (and quadratic) age term as exogenous instrumental variables. We found the corresponding estimates for a linear and quadratic control function to be very similar to the two-stage estimates discussed above.

In assessing the credibility of our RD estimates, it is also important to reconsider the validity of the underlying continuity conditions for our application. The assumptions that $E[u|A]$ and $E[\alpha_i|A]$ are continuous functions of the applicant’s age at the time of the award decision formalize the notion that observations in a small interval left and right of the cutoff should have comparable outcomes both given treatment and without treatment (as in a randomized assignment). As the first four stages of the disability determination process do not involve the grid, applicants reaching the fifth stage with ages just below and above each cutoff can be expected to be comparable in all characteristics, including their health status.

However, if initial applicants were aware of the existence of the grid, this could lead to violations of these comparability assumptions, as applicants may change their behavior (misreport their age, or time their application decision) to increase their likelihood of being accepted into the program. In this case, applicants who apply soon after they turn 55 may systematically differ from those who apply at age 54, thereby compromising the validity of our approach. However, given that the rule, while commonly used by DI administrators in evaluating disability cases, is only used for internal purposes and not advertised, we do not believe this to be a problem in our case.

It is also possible that for reasons unrelated to the disability determination process the composition of the applicant pool changes in a discontinuous way at the cutoff point. For example, if many individuals become eligible for early retirement benefits at age 55, and if retirees are more likely to apply for DI benefits, this could lead to a sudden increase and change in the composition of the applicant pool at age 55, potentially compromising RD’s continuity assumptions. While there is little evidence of a large jump at age 55 in the retirement rate for the total labor force, there is evidence of an increase for federal employees

²¹These estimates, as well as those based on the IV procedure which will be discussed next, can be found in the working paper version of this article.

who, with sufficient work experience, become entitled to early retirement benefits at age 55 (Asch, Haider and Zissimopoulos, 2005). However, to see a sudden change in the composition in the applicant pool at age 55, one would generally expect to find a jump in the application rate at that age. Figure 6 which plots the distribution of applicants in our sample against the age of the award decision date shows no evidence of heaping near either cutoff age. While this evidence provides little indication of a sudden change in the composition of the applicant pool near the age 55 cutoff, it is important to point out that an increase in the proportion of applicants eligible for early retirement benefits at age 55, would imply that the lower labor supply observed several years later for this group could in part be due to the receipt of pension benefits and would therefore cause our estimate to be an over-estimate of the true DI effect.

One way to analyze the comparability of applicants with ages just above and below an age cutoff at the time of disability determination is to test whether they differ in average observed characteristics. Accordingly we performed simple t-tests, as well as local Wald estimates which instead of labor supply, take applicants' observed characteristics as dependent variables. In addition, we estimated the discontinuity at age 55 in a piecewise linear specification of the applicant's characteristic as a function of age, and computed the F-statistic of the joint significance of any jumps at the three age cutoffs. While, depending on the width of the age interval, we find the difference in the proportion of applicants who completed high school to be sometimes positive and sometimes negative, the difference is never statistically significant. Local Wald estimates and the F-statistics similarly indicate that the difference in this proportion, between those a little older and younger than the age 55 cutoff or at all three age cutoffs combined, is never statistically significant. The same is true when we consider the applicant's post-high school education, sex and race. While we find that applicants with ages just above 55 generally have lower college enrollment rates, and are more likely to be female, these differences are never statistically significant. Moreover, as one would expect these differences to be associated with lower labor supply levels, if anything this would mean that we are over-estimating the decline in employment and hours worked associated with disability receipt.

While this evidence enhances the credibility of the RD approach, it is important to recall that our RD estimates are likely to be biased downward due to misclassification errors in our benefit award indicator, which ignores award decisions made above beyond the reconsideration stage. To assess the magnitude of this bias it is important to first consider whether the rate at which rejected applications are successfully appealed is continuous at the cutoff age. This would generally require the rate of appeal, and subsequent award rate among those rejected at the reconsideration stage to be continuous in the persons age at the award decision date. As we found individuals with ages just below and above the cutoff to be comparable in observed characteristics, it seems reasonable to expect their rate of appeals if rejected to be

comparable as well. Moreover, it is important to note that the ALJ and appeal councils do not make use of the grid in their decision-making, suggesting that the ultimate success rate of these appeals would be comparable as well. Unfortunately, lack of award data beyond the reconsideration prevents us from directly testing these conditions.

Under the assumption of a continuous successful appeal rate, the actual (rather than observed) disability benefit award rate will have a discontinuity at each cutoff age that will be reduced proportionally at the rate at which rejected applications are successfully appealed. Based on the computations in section 4, with 24% of rejected applicants in our sample expected to be misclassified DI beneficiaries, we estimate the discontinuities in the true award rate to be approximately 76% the size of the estimated discontinuities in the observed award rate, indicating that our RD estimates of the ratio (7) underestimate the true work disincentive effects of DI benefit receipt by 32%. This translates into corrected point estimates of -27 (-8) for the short-run (long-run) CFE and -8 (-6) and -14 (-11), for the 2-year and 3-year short-run (long-run) Local Wald estimates, which together with similarly corrected estimates for monthly hours are reported in Table 5 in italics. Similarly to our analysis based on the comparison group approach, we can also compute an upper bound on the misclassification bias, by assuming that all those rejected at the reconsideration stage in our sample went on to make successful appeals. As 47 percent of rejected applicants in our sample did not request a reconsideration, this group constitutes at most 53 percent of our sample. With the discontinuity in the true award rate being approximately 47% the size of the estimated discontinuity in the award rate, this implies corrected short-run (long-run) control function estimates of -43 (-13) and 2-year short-run (long-run) local Wald estimates of -12 (-10), respectively.²²

While our misclassification bias adjusted RD estimates are significantly larger, they, and the similarly corrected comparison group based estimates discussed in section 4 are considerably smaller (in absolute value) than the estimates reported by Bound (1989) for the 1970s. As discussed earlier, there are important differences between our sample and that used by Bound which may account for this result. While Bound's sample only included older male applicants aged 45-64, our sample includes both male and female applicants aged 35 to

²²What if the rate of successful appeals were not continuous, but instead contained jumps at the cutoffs? In case of a positive jump this would cause our estimates to be over-estimates (in absolute sense) of DI's work disincentive effect. To assess the case of a negative jump, consider the case where instead of being a comparable 24% the successful appeals rate for those just below the cutoff was, for example, double that of those above the cutoff age, say 30% and 15% respectively. With observed award rates of 0.41 and 0.82 (0.49 and 0.82) for the short-run 2-year (3-year) samples just below and above the age 55 cutoff age, then the estimated discontinuity in the true award rate at age 55 for the short-run sample would be reduced from 0.33 to 0.20 based on the 2-year age interval, and from 0.41 to 0.26 based on the 3-year interval. This would imply that the true effect of DI benefit receipt on labor force participation would be approximately 65% (58%) larger than estimated here. The corrected local Wald estimates in this case indicate that approximately 10 percent (16 percent) of DI applicants would have worked without DI benefits.

61. Our sample also includes individuals who applied only to the SSI program, a group that constitutes 28% of our total sample. Table 6 investigates the role of sample composition by reporting simple OLS estimates of the difference in the labor force participation rate of DI beneficiaries and rejected applicants, as well as RD estimates of the labor force participation effect of DI benefit receipt for several subgroups of our long-run sample.²³ The OLS estimates and RD estimates for the full sample correspond to the bias-adjusted estimates reported earlier in Tables 2 and 5.

Shown in column 2 of the table are estimates based on a sample restricted to SSDI applicants aged 45 or more at the time of the award decision. As the sample excludes those ineligible for SSDI benefits due to insufficient work experience in covered employment, one may expect a lower labor force attachment for this group. Unlike Bound's sample, this sample includes both male and female applicants. As the latter group may be less responsive in their labor supply behavior to DI benefit receipt, we also show estimates in column 3 based on a sample of male SSDI applicants over age 45, a sample that is comparable to that used by Bound.

Focussing first on the simple comparison group estimates, we find these to be very stable across samples indicating that the difference between our results and that of Bound is not due to differences in sample composition. The RD estimates for the long-run sample indicate that the exclusion of SSI applicants from the sample leads to slightly larger effect estimates, but show little additional change when also female applicants are excluded. The overall magnitude of the long-run effect estimates remains small, indicating that the inclusion of SSI and female applicants in our sample does not explain the smaller labor supply effects. Finally, the RD estimates for the short-run sample are more sensitive to changes in sample composition, with estimates increasing in absolute magnitude after SSI and female applicants are excluded from the sample. The resulting estimates, even though not very precisely estimated, vary between a drop in the labor force participation rate of 23 and 40 percentage points depending on the RD estimation method and discontinuity sample used.

7 Conclusion

In this paper we assessed the work disincentive effect of the Disability Insurance program during the 1990s, based on a new data set in which administrative disability application and award records were merged with the 1990-1996 panels of the SIPP. Using a comparison group approach suggested by John Bound, we estimate that during the 1990s the labor force participation rate of DI beneficiaries would have been at most 20 percentage points higher

²³All estimates in the table have been adjusted upwards by 32% to account for the misclassification errors in award status.

had none received benefits. Similarly, we estimate that DI receipt leads to a reduction in average monthly hours of work of at most 30 hours.

We also apply a regression discontinuity approach to find even smaller labor supply responses for a group of ‘marginal’ applicants whose medical condition is more difficult to assess and whose disability determination is based on vocational factors. Based on our longer-term sample the RD estimates indicate that DI decreases labor supply by approximately 16 to 20 hours per month and participation by 6 to 12 percentage points. This applicant group represents a non-trivial proportion of applicants and of beneficiaries. Between 1980 and 1990 the number of applicants who qualified on the basis of vocational criteria increased from 26% to 37% (Lahiri et al., 1995), and it represents 39% of all applicants in our sample. The RD estimates measure the average labor supply response for this group to a change in one component of the DI program – an age cutoff in the medical vocational grid.

In terms of spending, the Disability Insurance program is one of the largest social insurance programs in the United States. In 1999, over 100 billion dollars was spent providing medical benefits and cash payments to beneficiaries and their families. Credible estimates or bounds for the effect of DI on labor supply are therefore extremely important to policymakers. Increasing the age eligibility cutoff values would be one way to control the growing costs and caseload of the DI program that is likely to accompany the Social Security Administration’s increase in the normal retirement age. The regression discontinuity estimates presented in this paper indicate that the impact of a small increase in the age cutoffs would only have a modest overall effect on labor force participation. Our estimates also indicate that within this group responses vary with individual characteristics, with male and SSDI applicants generally showing somewhat larger labor supply responses than female and SSI applicants.

Combined our findings suggest that during the 1990s the work disincentive effects of the disability insurance program were rather modest: a large majority of applicants would not have worked even if none had received disability benefits. At first sight these findings appear to be at odds with time series evidence presented by Bound and Waidmann (2002) of a close association during the 1990s between the fraction of the working aged population receiving DI benefits and the growth in the fraction of the population that identifies itself as health-limited and out of work. These trends would be consistent with a movement of men and women in relative poor health out of the labor force and onto disability roles, suggesting that those drawn to apply for disability benefits when the program was expanding during the 1990s would have worked had they not applied for, and in many cases, been awarded DI benefits.

However, the results shown by Bound and Waidmann are inconclusive about the causality of the observed trends. In addition to an increase in generosity and a relaxation in the requirements to qualify for DI benefits, there are several other factors which are likely to

have contributed to the observed employment decline. Several researchers have attributed some of the decline in employment among the disabled to the introduction of the Americans with Disability Act of 1990 which took effect in 1992 and which led to an increase in the cost to employers of hiring such workers (DeLeire 2000, 2003; Acemoglu and Angrist 2001; Jolls and Prescott 2004).

There is also evidence pointing to the more general role of declining labor market conditions for low skilled workers during the 1990s (Juhn et al., 2002). Decreased opportunities for low-skilled workers during the early and middle part of the decade, coupled with a decreasing real wage and increasing real disability benefit amounts for this segment of the labor force is likely to have induced some employed and laid off workers, to apply for disability benefits. Paired with a liberalization of screening standards, the drop in demand for older, less skilled workers in poor health may therefore have led to higher DI reciprocity rates (Rupp and Stapleton, 1995). Our finding of a small work disincentive effect associated with disability benefit receipt, even smaller than those found for the 1970s, is therefore likely to reflect the unfavorable labor market conditions, and the lower labor force attachment of applicants to the DI program during this period. During the period between the 1970s (when Bound's data was collected) and the nineties, there has also been a significant increase in the labor supply of married women. If the presence of these women creates a source of income maintenance then this may also have contributed to the smaller employment effect associated with benefit receipt.

In conclusion, our results suggest that most of those drawn to the DI program during the 1990s would not have worked in absence of disability benefits, but instead would have been unemployed or on other types of welfare support. Whether they would not work because of their medical conditions, unfavorable labor market conditions, or increased spousal labor supply is beyond the scope of this paper, but is an important area for further inquiry.

Acknowledgements

We have benefited from helpful comments from John Bound, John DiNardo, Donald Parsons, Guido Imbens, Thomas Lemieux, Jack Porter, Josh Angrist, Bob Willis, Denton Vaughan and seminar participants at Duke, Purdue, Syracuse, University of Michigan, and the 2003 Regression-Discontinuity Conference in Banff. We would like to acknowledge grant support from the Demography and Economics of Aging Research Program at UNC from the NIA (grant number NIA P30 AG024376) and from the Northwestern University-University of Chicago Joint Center for Poverty Research through a Census Bureau Research Development Grant for Policy Research on U.S. Poverty. We are grateful to the Research Triangle Census Data Center for providing us access to the data, and especially to the Triangle RDC administrator Kirk White for his exceptional support. Support for this research at the Triangle RDC from the NSF (award no. ITR-0427889) is also gratefully acknowledged. The research in this paper was conducted while the authors were Special Sworn Status researchers of the U.S. Census Bureau at the Triangle Census Research Data Center. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the Census Bureau, nor do the views and opinions expressed in the paper reflect those of the Federal Reserve Bank of New York or the Federal Reserve System as a whole. This paper has been screened to insure that no confidential data are revealed.

Bibliography

- Acemoglu, D. and J. D. Angrist, 2001, Consequences of Employment Protection? The Case of the Americans with Disabilities Act, *Journal of Political Economy* 109, 915-957.
- Angrist, J. D. and V. Lavy, 1999, Using Maimonides' Rule to Estimate the Effect of Class Size on Student Achievement, *Quarterly Journal of Economics* 114(2), 533-575.
- Asch, B. J., S. J. Haider and J. Zissimopoulos, 2005, Financial Incentives and Retirement: Evidence from Federal Civil Service Workers, *Journal of Public Economics* 89(2-3), 427-440.
- Autor, D. H. and M. G. Duggan, 2003, The Rise in the Disability Rolls and the Decline in Unemployment, *Quarterly Journal of Economics* 118(1), 157-206.
- Bound, J., 1989, The Health and Earnings of Rejected Disability Insurance Applicants, *American Economic Review* 79 (3), 482-503.
- Bound, J., 1991, The Health and Earnings of Rejected Disability Insurance Applicants: Reply, *American Economic Review* 81(5), 1427-1434.
- Bound, J. and R. V. Burkhauser, 1999, Economic analysis of transfer programs targeted on people with disabilities, in: O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3C (North-Holland, Amsterdam).
- Bound, J., R. V. Burkhauser and A. Nichols, 2003, Tracking the Household Income of SSDI and SSI Applicants, in: S. Polachek, ed., *Research in Labor Economics* 22, (JAI, Elsevier Science), 113-158.
- Bound, J. and T. Waidmann, 1992, Disability Transfers, Self-Reported Health, and the Labor Force Attachment of Older Men: Evidence from the Historical Record, *Quarterly Journal of Economics*, 107(4), 1393-419.
- Bound, J. and T. Waidmann, 2002, Accounting for Recent Declines in Employment Rates among Working-Aged Men and Women with Disabilities, *Journal of Human Resources* 37(2), 231-250.
- Burtless, G., 1999, An Economic View of Retirement, in: H.J. Aaron ed., *Behavioural Dimensions of Retirement Economics* (Brookings Institution Press Washington DC and Russell Sage Foundation New York).
- DeLeire, T., 2000, The Wage and Employment Effects of the Americans with Disabilities Act, *Journal of Human Resources* 35(4), 693-715.
- Chen, S. and W. van der Klaauw, 2005, The Work Disincentive Effects of the Disability Insurance Program in the 1990s, unpublished manuscript, University of North Carolina at Chapel Hill.
- DeLeire, T., 2003, The Americans with Disabilities Act and the Employment of People with Disabilities, in: D. C. Stapleton and R. V. Burkhauser (eds), *The Declining Employment of People with Disabilities: A Policy Puzzle* Kalamazoo (W.E. Upjohn Institute: MI), 259-275.

- Gruber, J., 2001, Disability Insurance Benefits and Labor Supply, *Journal of Political Economy* 108(6), 1162-1183.
- Hahn, J., P. Todd and W. van der Klaauw, 2001, Identification and Estimation of Treatment Effects with a Regression- Discontinuity Design, *Econometrica* 69(1), 201-209.
- Haveman, R. H. and B. L. Wolfe, 1984a, The Decline in Male Labor Force Participation: Comment, *Journal of Political Economy* 92(3), 532-541.
- Haveman, R. H. and B. L. Wolfe, 1984b, Disability Transfers and Early Retirement: A Causal Relationship? *Journal of Public Economics* 24, 47-66.
- Heckman, J. J. and Richard Robb, 1985, Alternative Methods for Evaluating the Impacts of Interventions, in: J. J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*, (Cambridge University Press, New York), 156-245.
- Hu, J., K. Lahiri, D. R. Vaughan, and B. Wixon, 2001, A Structural Model of Social Security's Disability Determination Process, *The Review of Economics and Statistics* 83 (2), 348-361.
- Imbens, G. W. and J. D. Angrist, 1994, Identification and Estimation of Local Average Treatment Effects, *Econometrica* 62, 467-476.
- Jolls, C. and J. J. Prescott, 2004, Disaggregating Employment Protection: The Case of Disability Discrimination, NBER Working Paper 10740.
- Juhn, C., K. M. Murphy and R. H. Topel, 2002, Current Employment, Historically Contemplated, *Brookings Papers on Economic Activity* 1, 79-116.
- Lahiri, K., D. R. Vaughan, and B. Wixon, 1995, Modelling SSA's Sequential Disability Determination Process Using Matched SIPP Data, *Social Security Bulletin* Vol. 58.
- Leonard, J., 1986, Labor Supply Incentives and Disincentives for Disabled Persons, in: M. Berkowitz and M. A. Hill, eds., *Disability and the Labor Market: Economic Problems, Policies and Programs*, (Utica, NY: Industrial and Labor Relations Press).
- Newey, W. K., J. L. Powell, and J. R. Walker, 1990, Semiparametric Estimation of Selection Models: Some Empirical Results, *American Economic Review* 80, 324-28.
- Parsons, D. O., 1980, The Decline in Male Labor Force Participation, *Journal of Political Economy* 88(1), 117-134.
- Parsons, D. O., 1991, The Health and Earnings of Rejected Disability Insurance Applicants: Comment, *American Economic Review* 81(5), 1419-1426.
- Porter, J., 2003, Estimation in the Regression Discontinuity Model, unpublished manuscript, Harvard University.
- Rupp, K. and D. C. Stapleton, 1998, Growth in Disability Benefits. Explanations and Policy Implications, eds., *Social Security Administration and The Lewin Group*.
- Slade, F. P., 1984, Older Men: Disability Insurance and the Incentive to Work, *Industrial Relations* 23(2), 260-69.
- Thistlethwaite, D. L. and D. T. Campbell, 1960, Regression-Discontinuity Analysis: An

alternative to the ex post facto experiment, *Journal of Educational Psychology* 51, 309-317.

U.S. House of Representatives, 1993, Committee on Ways and Means, 1993 Greenbook: Overview of Entitlement Programs, U.S. Government Printing Office Washington D.C.

U.S. House of Representatives, 1996, Committee on Ways and Means, 1996 Greenbook: Overview of Entitlement Programs, U.S. Government Printing Office Washington D.C.

U.S. House of Representatives, 2000, Committee on Ways and Means, 2000 Greenbook: Overview of Entitlement Programs, U.S. Government Printing Office, Washington D.C.

U.S. Social Security Administration, Social Security Bulletin Annual Statistical Supplement, various years.

Van der Klaauw, W., 2002, Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach, *International Economic Review* 43(4), 1249-1287.

TABLE 1: SAMPLE STATISTICS

Variable	All Applicants		Stage 5 Applicants	
	Short Run	Long Run	Short Run	Long Run
Age at Award Decision	49.5	49.1	49.7	49.3
High School (%)	35.9	35.1	37.5	36.5
More than High School (%)	26.4	26.3	23.2	23.6
Non-white (%)	30.7	30.7	29.0	29.4
Female (%)	51.7	52.0	46.5	47.4
Labor Force Partipation (%)	14.3	14.9	13.1	13.9
Monthly Hours Worked	20.4	21.9	18.7	20.0
Weekly Hours if Employed	32.9	33.9	32.9	33.2
Months Since Award Date	17.6	36.7	17.5	36.2
Observations	3575	4467	1386	1723

Short run: between 1 and 24 months after date of award decision.

Long run: between 1 months and 11 years after date of award decision.

Age is measured at the date of the award decision. In 66% of cases this corresponds to the initial application phase, for 34% it is at the reconsideration stage.

Monthly hours of work and **LFP** are measured in the last SIPP survey month available for the individual at which his/her age was less or equal to 61.

TABLE 2: STATISTICS BY AWARD STATUS AND COMPARISON GROUP ESTIMATES

Variable	All Applicants						Stage 5 Applicants					
	Short Run			Long Run			Short Run			Long Run		
	T=0	T=1	t	T=0	T=1	t	T=0	T=1	t	T=0	T=1	t
Age at Award Decision	48.0	51.1	-11.8	47.7	50.5	-12.6	46.4	53.0	-17.3	46.1	52.6	-19.1
HS (%)	37.1	34.5	1.6	36.8	33.2	2.5	41.1	34.0	2.8	39.6	33.3	2.7
More than HS (%)	25.8	27.1	-0.9	25.3	27.4	-1.6	23.3	23.1	0.1	23.5	23.6	-0.0
Non-white (%)	34.5	26.7	5.1	34.2	27.0	5.2	31.8	26.2	2.3	32.5	26.2	2.8
Female (%)	54.8	48.4	3.9	54.6	49.2	3.6	45.9	47.0	-0.4	46.5	48.4	-0.8
LFP (%)	21.2	6.9	12.5	22.2	7.1	14.5	20.5	5.9	8.2	21.5	5.8	9.6
Monthly Hours Worked	30.6	9.6	11.5	33.0	9.9	13.7	29.8	7.8	7.8	32.3	7.0	9.7
Weekly Hours ¹	33.3	32.2	0.8	34.3	32.4	1.5	33.6	30.6	1.1	34.7	27.7	3.0
Months Since Decision	17.9	17.3	2.2	37.8	35.6	2.6	18.0	17.2	2.0	39.1	33.3	4.4
Obs	1844	1731		2318	2149		688	698		884	839	
<u>Unconditional</u>												
LFP Effect (std error)	-14.33		(1.15)	-15.10		(1.04)	-14.62		(1.77)	-15.65		(1.62)
Hours Effect (std error)	-21.05		(1.83)	-23.05		(1.68)	-22.02		(2.83)	-25.33		(2.60)
<u>Conditional</u>												
LFP Effect (std error)	-12.93		(1.17)	-13.27		(1.05)	-11.16		(1.95)	-11.07		(1.78)
Hours Effect (std error)	-18.81		(1.86)	-20.19		(1.70)	-16.43		(3.11)	-18.35		(2.85)
<u>Bias Adjusted</u>												
LFP Effect (std error)	<i>-18.92</i>		(1.52)	<i>-19.93</i>		(1.37)	<i>-19.30</i>		(2.34)	<i>-20.66</i>		(2.14)
Hours Effect (std error)	<i>-27.79</i>		(2.42)	<i>-30.43</i>		(2.22)	<i>-29.07</i>		(3.74)	<i>-33.44</i>		(3.43)

¹ Average weekly hours for those with positive monthly hours.

Beneficiaries are indicated by $T = 1$ while rejected applicants are denoted by $T = 0$. The t-statistics test for equality of average characteristics between beneficiaries and rejected applicants. The conditional labor supply estimates were obtained from OLS regressions on benefit receipt, age, and education, race and sex dummies. The reported bias corrected estimates were obtained by multiplying the unconditional estimates in the table by 1.32 to correct for misclassification errors in our benefit receipt variable.

TABLE 3: SEGMENT OF THE MEDICAL VOCATIONAL GRID

7th Through 11th Grade		
Residual Functional Capacity	Age	(Semi) Skilled/Non-transferable
SEDENTARY	60-64	D
	55-59	D
	50-54*	D
	45-49	N
	18-44	N
LIGHT	60-64	D
	55-59*	D
	50-54	N
	45-49	N
	18-44	N
MEDIUM	60-64	N
	55-59	N
	50-54	N
	45-49	N
	18-44	N

(D=Disabled and N=Not Disabled)

TABLE 4: ESTIMATES OF AWARD EQUATION

Coefficient Estimate	Sample		
	Extended	Short Run	Long Run
γ_1	-0.016 (0.040)	-0.013 (0.074)	0.016 (0.065)
γ_2	0.093 (0.045)	0.119 (0.082)	0.074 (0.071)
γ_3	0.287 (0.037)	0.313 (0.068)	0.217 (0.060)
ψ_{00}	0.394 (0.162)	0.456 (0.285)	0.511 (0.249)
ψ_{01}	-0.003 (0.003)	-0.003 (0.006)	-0.005 (0.005)
ψ_{11}	0.073 (0.099)	0.004 (0.184)	0.027 (0.160)
ψ_{12}	0.239 (0.131)	0.244 (0.241)	0.422 (0.208)
ψ_{13}	-0.213 (0.097)	-0.160 (0.186)	-0.305 (0.164)
Obs	4294	1386	1723

Linear probability model estimates of award equation specified as $E[T_i|A] = g(A_i) + \sum_{j=1}^3 \gamma_j I\{A_i \geq \bar{A}_j\}$ where $g(A) = \psi_{00} + \psi_{01}A + \sum_{j=1}^3 \psi_{1j}(A - \bar{A}_j)1\{A_i \geq \bar{A}_j\}$. Heteroskedasticity corrected standard errors in parentheses. The estimation samples only includes stage 5 applicants.

TABLE 5: RD ESTIMATES OF LABOR SUPPLY EFFECTS

Sample	CFE		Local Wald at age 55					
			± 2 yr interval		± 3 yr interval		± 4 yr interval	
Labor Force Participation								
Short Run	-20.1	(7.0)	-5.8	(8.9)	-10.4	(6.0)	-10.7	(5.7)
	<i>-26.5</i>	<i>(9.2)</i>	<i>-7.7</i>	<i>(11.7)</i>	<i>-13.7</i>	<i>(7.9)</i>	<i>-14.1</i>	<i>(7.5)</i>
Obs	1386		268		374		490	
Long Run	-5.9	(6.5)	-4.8	(11.4)	-8.3	(6.6)	-8.8	(5.9)
	<i>-7.8</i>	<i>(8.6)</i>	<i>-6.3</i>	<i>(15.0)</i>	<i>-11.0</i>	<i>(8.7)</i>	<i>-11.6</i>	<i>(7.8)</i>
Obs	1723		332		463		613	
Monthly Hours of Work								
Short Run	-31.9	(11.5)	-15.6	(14.5)	-16.8	(9.2)	-17.8	(8.4)
	<i>-42.1</i>	<i>(15.2)</i>	<i>-20.6</i>	<i>(19.1)</i>	<i>-22.2</i>	<i>(12.1)</i>	<i>-23.5</i>	<i>(11.1)</i>
Obs	1386		268		374		490	
Long Run	-12.1	(10.4)	-14.1	(17.2)	-15.4	(10.1)	-15.0	(8.5)
	<i>-16.0</i>	<i>(13.7)</i>	<i>-18.6</i>	<i>(22.7)</i>	<i>-20.3</i>	<i>(13.3)</i>	<i>-19.8</i>	<i>(11.2)</i>
Obs	1723		332		463		613	
First Stage Discontinuity Estimates								
Short Run	0.31	(0.07)	0.33	(0.05)	0.41	(0.05)	0.38	(0.04)
Long Run	0.22	(0.06)	0.26	(0.05)	0.34	(0.04)	0.34	(0.04)

Estimates based on short and long run samples of stage 5 applicants. Standard errors in parentheses, with those reported for the Local Wald estimates being bootstrap standard errors. Bias-adjusted estimates shown in italics are based on a misclassification bias correction factor of 1.32. The optimal polynomial specification of $k(A)$ for the two-stage control function estimates (CFE) was linear, for both the short run and long run samples and for labor force participation and monthly hours as dependent variable. First-stage estimates reported for the CFE estimator are for the discontinuity at age 55 only. See Table 4 for estimates of the discontinuities at ages 45 and 50.

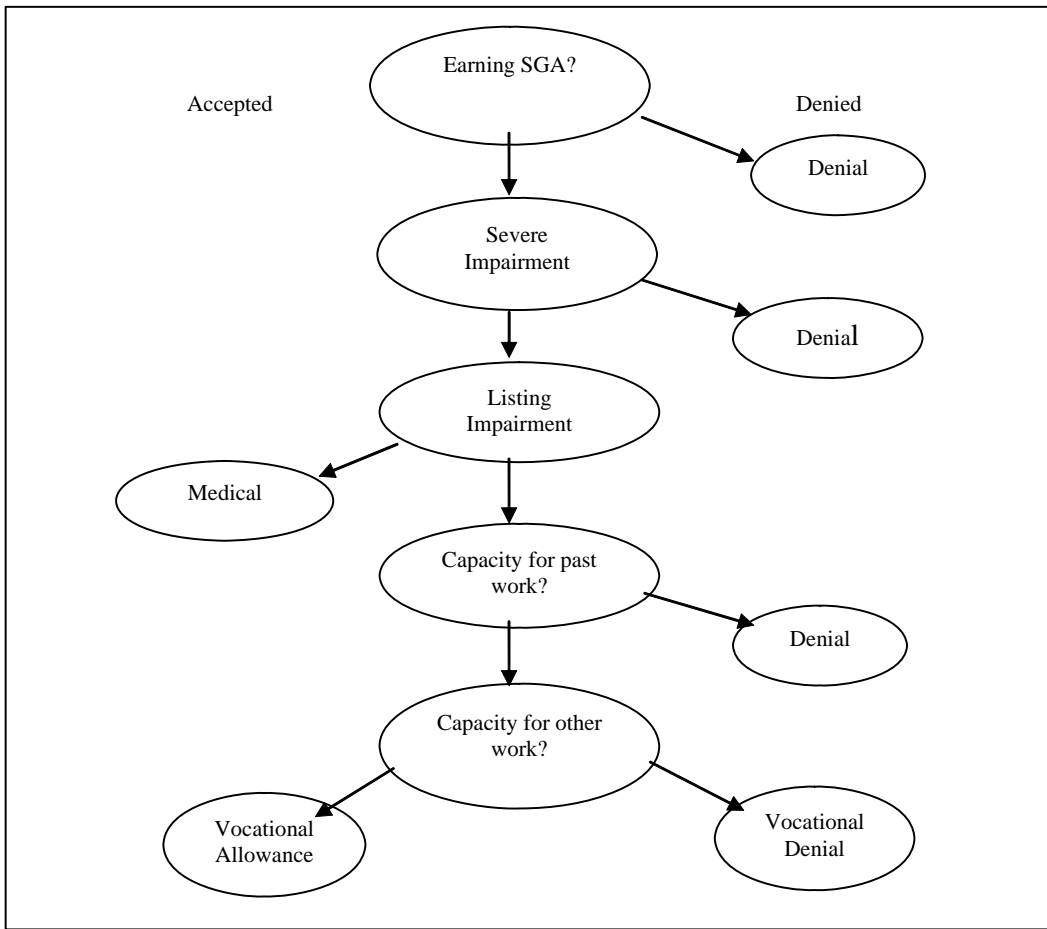
TABLE 6: LFP ESTIMATES FOR ALTERNATIVE SUBSAMPLES

Sample	Estimate	Total Sample	Age \geq 45 SSDI applicants	Male, Age \geq 45 SSDI applicants
Short Run				
All Applicants	OLS	-18.9 (1.5) [3575]	-16.1 (2.0) [1811]	-18.1 (2.6) [999]
Stage 5 Applicants	CFE	-26.5 (9.2) [1386]	-28.4 (21.1) [672]	-40.0 (34.2) [430]
	\pm 2yr Local Wald	-7.7 (11.7) [268]	-8.2 (13.9) [191]	-23.0 (21.9) [112]
	\pm 3yr Local Wald	-13.7 (7.9) [374]	-18.6 (9.9) [272]	-30.0 (14.7) [169]
	\pm 4yr Local Wald	-14.1 (7.5) [490]	-22.0 (10.0) [346]	-33.1 (14.0) [216]
Long Run				
All Applicants	OLS	-19.9 (1.4) [4467]	-16.0 (1.8) [2189]	-18.0 (2.4) [1192]
Stage 5 Applicants	CFE	-7.8 (8.6) [1723]	-29.3 (21.6) [804]	-18.0 (34.2) [506]
	\pm 2yr Local Wald	-6.3 (15.0) [332]	-8.1 (19.0) [229]	-5.4 (25.5) [136]
	\pm 3yr Local Wald	-11.0 (8.7) [463]	-16.6 (11.6) [324]	-16.0 (15.3) [202]
	\pm 4yr Local Wald	-11.6 (7.8) [613]	-20.6 (10.6) [420]	-24.4 (14.3) [260]

All estimates and standard errors have been bias-adjusted through multiplying by 1.32 to correct for measurement errors in the DI benefit receipt variable. Standard errors are reported in parentheses, which represent bootstrap standard errors for the Local Wald estimates, and

heteroskedasticity consistent standard errors that have been corrected for generated regressors for the CFEs. The number of observations used for each sample and estimator are reported in brackets.

Figure 1: The 5 Stages of the Social Security Disability Determination Process



(Source: Lahiri et al, 1995).

Figure 2: DI Award Rate
Extended Sample

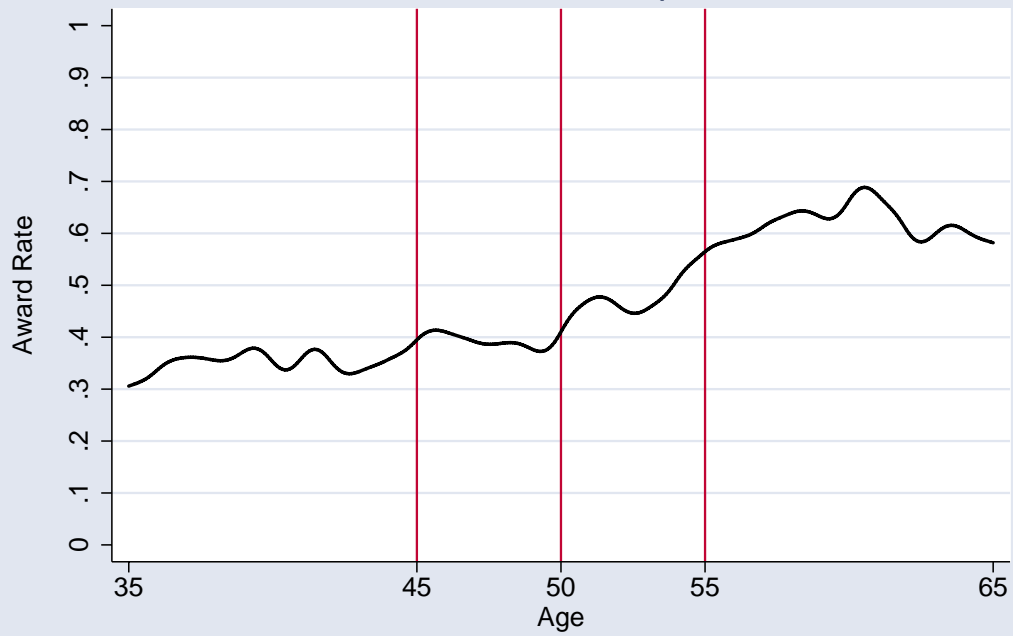


Figure 3: DI Award Rate By Determination Stage
Extended Sample

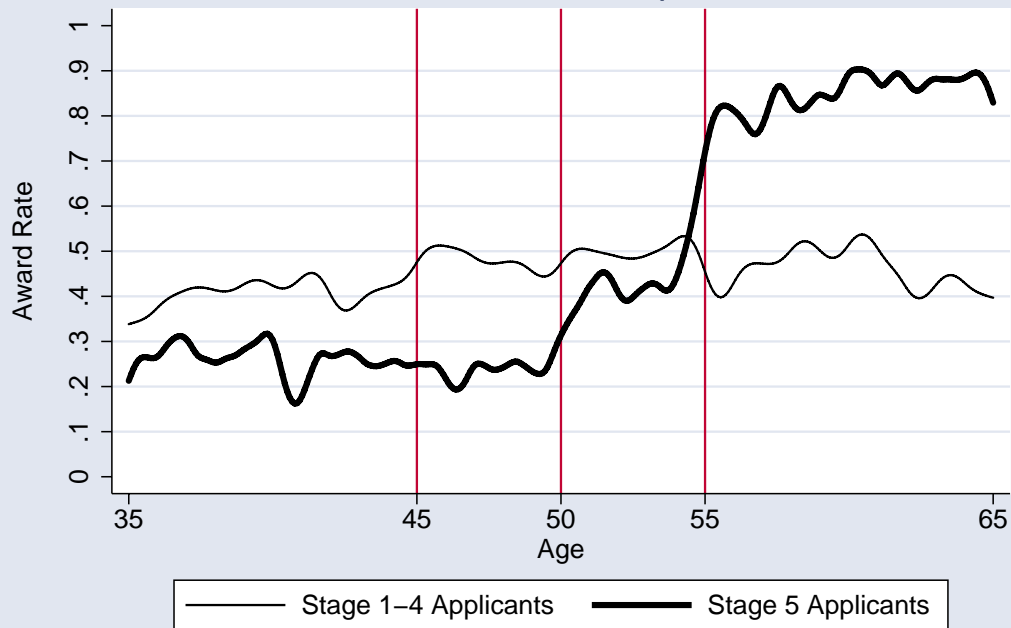


Figure 4: DI Award Rate For Stage 5 Applicants
Extended Sample

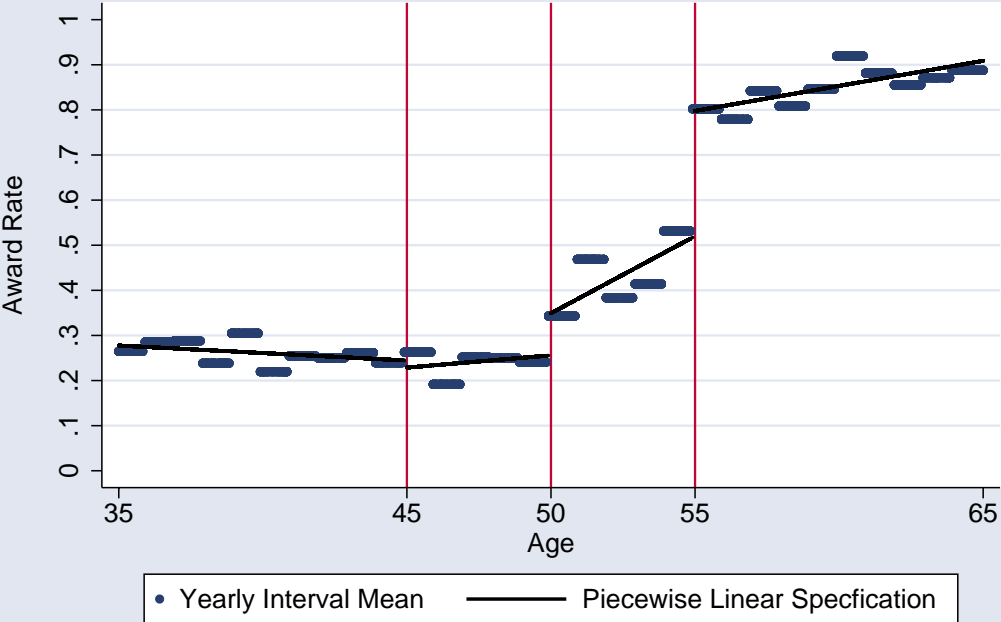


Figure 5: LFP Rate For Stage 5 Applicants
Short Run Sample

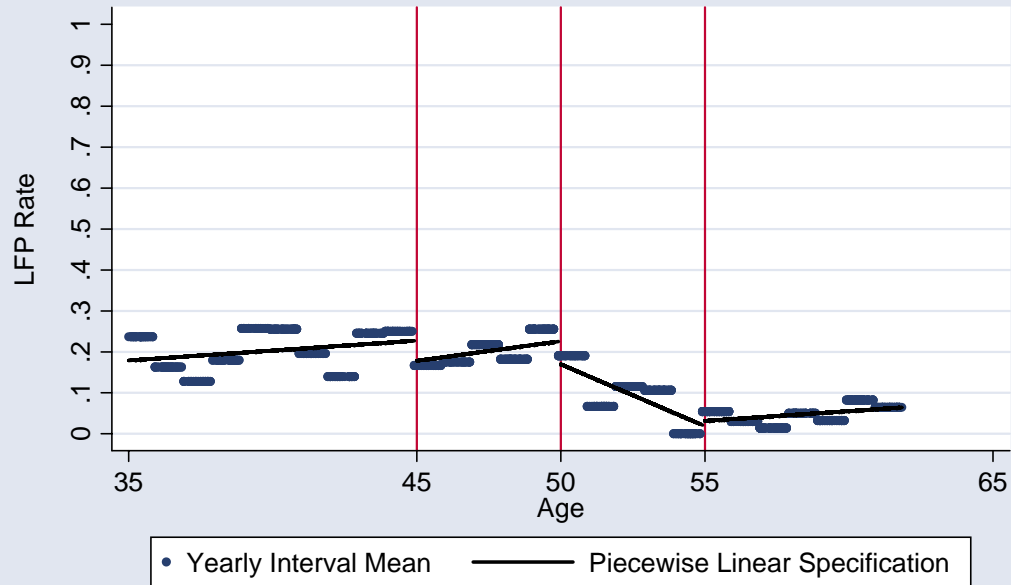
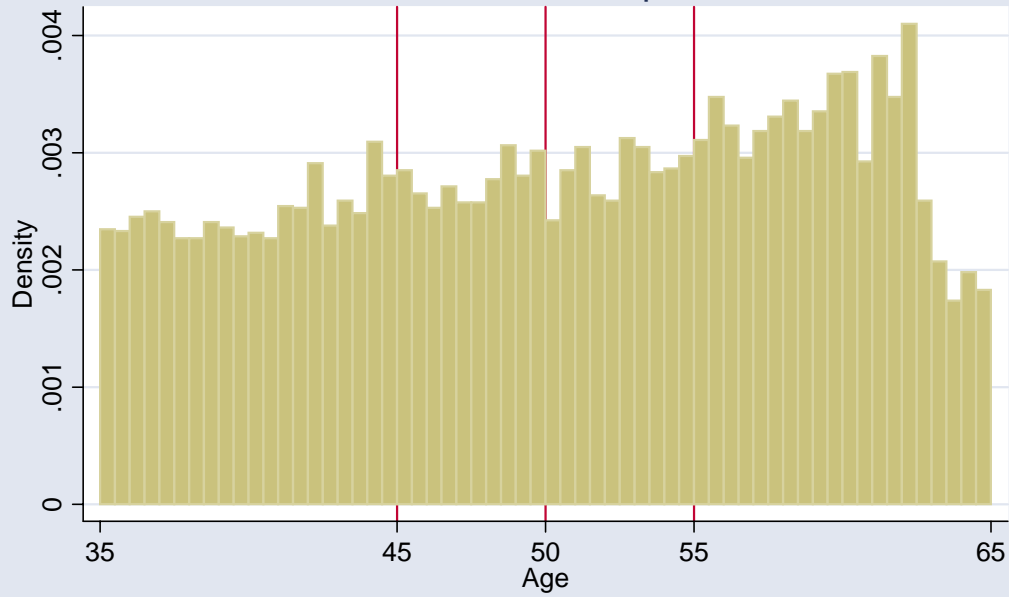


Figure 6: Distribution of Age At Time of Decision
Extended Sample



Note: 6 month intervals

Table A1: Medical Vocational Grid

Illiterate or No Fluency in English				1-6th Grade		7th Through 11th Grade		High School or Above	
Residual Functional Capacity	Age	Unskilled/None	(Semi) Skilled/Non-transferable	Unskilled/None	(Semi) Skilled/Non-transferable	Unskilled/None	(Semi) Skilled/Non-transferable	Unskilled/None	(Semi) Skilled/Non-transferable
	60-64	D	D	D	D	D	D	D	D
	55-59	D	D	D	D	D	D	D	D
SEDENTARY	50-54	D	D	* D	* D	* D	* D	* D	* D
	45-49	D	* N	N	N	N	N	N	N
	18-45	N	N	N	N	N	N	N	N
	60-64	D	D	D	D	D	D	D	D
	55-59	D	D	D	* D	* D	* D	* D	* D
LIGHT	50-54	D	* D	* N	N	N	N	N	N
	45-49	N	N	N	N	N	N	N	N
	18-45	N	N	N	N	N	N	N	N
	60-64	D	N	D	N	D	N	N	N
	55-59	D	* N	N(D)	N	N(D)	N	N	N
MEDIUM	50-54	N	N	N	N	N	N	N	N
	45-49	N	N	N	N	N	N	N	N
	18-45	N	N	N	N	N	N	N	N

(D=Disabled, N=Not Disabled)