# The Work Disincentive Effects of The Disability Insurance Program in the 1990s[1]


Susan Chen
Duke University[2]

and

H. Wilbert van der Klaauw
University of North Carolina at Chapel Hill[3]


August 2004

## Abstract


In this paper we evaluate the work disincentive effects of the Disability Insurance program during the 1990s. To accomplish this we construct a new large data set with detailed information on DI application and award decisions and use two different econometric evaluation methods. First, we apply a comparison group approach proposed by John Bound to estimate an upper bound for the work disincentive effect of the current DI program. Second, we adopt a Regression-Discontinuity approach that exploits a particular feature of the DI eligibility determination process to provide a credible point estimate of the impact of the DI program on labor supply for an important subset of DI applicants. Our estimates indicate that during the 1990s the labor force participation rate of DI applicants would have been at most 23 percentage points higher had none received benefits compared to the case where all 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.


---


[1]We have benefited from helpful comments from John Bound, John DiNardo, Jack Porter, Josh Angrist, Bob Willis, Denton Vaughan and participants at 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 and from the Northwestern University-University of Chicago Joint Center for Policy Research Census Bureau Research Development Grants for Policy Research on U.S. Poverty. We are grateful to the Research Triangle Census Data Center for providing us access to the data. The research in this paper was conducted while the authors were Census Bureau research associates at the Triangle Census Research Data Center. Research results and conclusions expressed are those of the author and do not necessarily indicate concurrence by the Census Bureau.



[2]Mailing address: Triangle Census Research Data Center, Department of Economics, 329E Social Sciences Building, Box 90097, Duke University, Durham, NC 27708-0097, Email: schen@econ.duke.edu.

[3]Mailing address: Department of Economics, Gardner Hall, CB#3305, UNC-Chapel Hill, Chapel Hill, NC 27599-3305. Email: vanderkl@email.unc.edu.

# 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 (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, little agreement exists 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. While previous studies of the effects of the DI program have generally found a negative correlation between male labor supply and benefit receipt, little attention has been given to the effect on the labor supply of female applicants. A deeper 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 managing 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 evaluator, this implies that DI receipt should be treated as endogenous in a regression 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. Compared to the earlier studies, they found much lower elasticity estimates of between 0.0006 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 adopt alternative ways to deal with the endogeneity of disability benefit receipt. Gruber (2000) exploits an exogenous policy change conducted in Canada in 1986 where the benefit levels of the rest of the country were adjusted upwards to meet those of Quebec province. He estimated 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 used differential time variation in average benefits across states 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 and the progressivity of the disability benefits formula provide an exogenous measure of program generosity independent of a workers underlying taste for work. They conclude that the disability system provided many of the low-skilled workers with a viable alternative to unemployment. They estimate that the unemployment rate in 1998 would have been 0.48 to 0.61 percentage points 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 levels and impacts of change in labor market conditions over time, which seems problematic since variation in the wage distribution over time across states can itself be expected to directly affect labor market outcomes.

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 at most half of DI beneficiaries would have worked 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, education level, and in their preferences (disutility) from working.

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 their LFP 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. In particular, 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 be 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 twenty 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 1995 the number of disabled individuals receiving DI increased by over 50 percent. This increase is particularly striking because, the Americans with Disabilities Act was passed 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 20 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 the control group evaluation approach described above 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 do not have one of most severe impairments and whose eligibility needs to be determined based on 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 the value of an individual's age with respect to given age cut-offs. 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 expand our analysis to study the employment effect of the DI program on the labor supply of both men and women. In addition, we go a step further and study the long term effects of DI benefit receipt on spousal employment.

## 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. It was designed to provide partial earnings replacement to all workers under age 65 and its benefit amounts 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.[1] In addition, more than 1.5 million individuals received SSDI benefits as dependent family members of disabled workers, of which over 90% were children. 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. 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 1999 this was earnings of less than $740 per month 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 about a quarter did so.

In 1999 the standard monthly SSI benefit to an individual was $467. 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. The standard benefit amount is capped if both husband and wife apply at $787. For children the average benefit amount is $489.

In order to be considered disabled an individual must first meet the eligibility rules and then meet the criteria laid out in the Code of Federal Regulations by the Social Security Agency (SSA). The Code requires that each applicant go through a sequential Disability

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

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 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 $740 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. At this stage the administrator also checks if the applicant has been completely detached from the work force for at least 5 months prior to the application date. 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 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 applicants 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 judgements. 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. In 1992 and 1995, respectively about 57 and 69 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 47 (49)% of the rejected applicants asked for reconsideration and about 17 (13)% of the people who applied for reconsideration were granted benefits.

Subsequent levels of appeal are no longer conducted at the State level. If an applicant is rejected at the reconsideration stage, they 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 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 an appeals court. 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 that appealed were granted benefits.

In summary about 14 (19)% of the initial applicant pool carried their claims to these higher levels in 1992 (1995), representing 26 (29)% of applications not awarded benefits at the state level.[2]

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).

[2]The sources for these statistics are the 1993 and 1996 Green Books.

# 3 Data

In this analysis we use data from a newly constructed data set that merges the 1990-1996 panels of the SIPP with restricted Social Security data containing detailed SSDI and SSI application and award information. The two 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 contains detailed information on each transaction that is made at each stage of the disability determination process excluding adjudications made above the State level. 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 subsequent appeals), so we define this date to be the date at which the final determination at the State level was made.[3]

While an individual's merged administrative data record may contain transactions data that precede and exceed 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-award 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 employment information is available. The first of these data sets has a short post-award time horizon and includes all individuals between

[3]The 831 file contains a nonneglible 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.

the age of 35 to 64 at the time of award for whom we have any employment information between 2 to 24 months after the award 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 64 at the time of award for whom we have any labor force information at least one month after the award decision. The time horizon ranges from 1 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 were pooled to form, what we call, the "Extended" sample.

The SIPP dataset provides information for 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 for this individual. Similarly, we define an individual to be a participant in the labor force if the number of hours worked in the last 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 defined earlier, measured in months. Another explanatory variable included in some of our analyses is an indicator for whether or not an applicant has a high school diploma. An individual was defined to have a high school education if he had 12 or more years of education or reported to have a high school diploma.

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. The average age of the short run sample is approximately 605 months (a little over 50 years old) and about 66% of the sample has at least a high school diploma. Measured on average 22 months after the award date, about 20% percent of the short run sample work and for those who work, the mean number of hours worked is 139 hours per month. The long run sample is on average a year older, is less likely to have a high school diploma, and has slightly lower labor force participation rates and mean hours of work measured on average 4 years after the award date.

When we only consider the subsamples of applicants whose award decisions were made on the basis of Stage 5 of the disability determination process, we find them to be surprisingly similar, the main difference being that the labor force participation rates for this group in both the short and long run samples are slightly lower. The similarily in employment rates suggest that while cases determined in stages 1 to 4 are more likely to include those with very bad an very good health, on average their employment rates and demographic characteristics are comparable to cases which reach stage 5. While the former group includes approximately equal groups of cases accepted and rejected on the basis of their medical conditions, the stage 5 cases represent the more difficult to decide (borderline) cases.[4] In our sample, the later group constitutes approximately 40% of all applications.

# 4 Results using Bound's Comparison Group Approach

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, Bound used matched social security earnings records to determine a rejected applicant's employment status, while we use self-reported employment from the SIPP. As social security records only account for individuals in covered jobs, those employed in some federal or state government jobs may not be counted as working. Our data includes data on employment hours and labor force participation for individuals in covered and non-covered employment. Third, and related to this, while Bound's original study was based only on information about SSDI applicants, our analysis also includes applicants to the SSI program. In fact, 44% of the applicants in our long-run sample applied only for SSI benefits, 28% applied only for SSDI benefits and the remaining 28% 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 (2002) 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 found that rejected SSI applicants were less likely to work than rejected SSDI

[4]A similar pattern was reported by Hu et al (1997) who found that approximately 33% of their total sample was refused benefits, and 30% was awarded benefits during stages 1 through 4.

applicants.[5] Finally, given the 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 to be evaluated by a local Social Security Administration office.

Fourth, instead of restricting our sample to male applicants, our data includes both male and female applicants. We would like to consider both separately because we recognize that the effect of DI on labor supply may differ for both groups, however, for efficiency reasons we wanted to keep the sample as large as possible. We have requested additional matched 831 files from the Census bureau and with a larger sample hope to study male and female applicants separately.

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. 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, the award date referes to the award date of the most recent application. In our data, for 65% of the cases the award date corresponded to the initial application phase, while for 35% of cases it corresponded to the award decision at the reconsideration stage. 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 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 the final award date. In addition, as shown in Table 5.1, the average time between the award date and the date at which the individual's employment status was measured was almost 2 years for the short run sample and 4 years for the long run sample. Finally, it is important to recall that these records exclude adjudications made above the State level. We return to the implications of these aspects of our data below.

[5] Bound et al use a dataset similar to ours but for different years; the 1990-1993 SIPP matched to the 831 file administered between 1977 to 1997.

# 5 Rejected Applicants As A Comparison Group

To better understand the two empirical approaches that we use in this study we begin with a discussion of the general evaluation problem. Our explanation below is outlined for individual labor supply and by extension, it also applies to spousal labor supply.

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 treatment is 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 in and out of the program. What we observe is,

$$y_i = t_i y_i(1) + (1 - t_i) y_i(0). \tag{1}$$

In our case, we want to evaluate the effect of disability benefit receipt $t$ on labor force participation $y$. Bound's proposed evaluation strategy for this evaluation problem is to use rejected applicants as a comparison group. To understand this approach, consider the evaluation of the average treatment effect on the treated. Estimation of a treatment effect in this case is hampered by the fact that we do not observe the percentage of DI beneficiaries that would have worked in the absence of the program, $E[y_i(0)|t_i = 1]$. As proposed by John Bound, we can however treat the observed employment rate of rejected DI applicants as an estimated upper bound, given that rejected applicants are generally healthier than accepted beneficiaries (Bound, 1989). 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. If, $E[y_i(0)|t_i = 1] \leq E[y_i(0)|t_i = 0]$ then,

$$E[\alpha_i|t_i = 1] \geq E[y_i(1)|t_i = 1] - E[y_i(0)|t_i = 0].$$

Note also that if, in addition, $E[y_i(1)|t_i = 0] \geq E[y_i(1)|t_i = 1]$, then the latter will represent an estimated bound on the average treatment effect $E[\alpha_i]$.

The validity of this 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 earlier, it is obvious that rejected applicants will on average be in better health than beneficiaries. 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 5.2 which shows rejected applicants on average to be younger and more educated. As we would expect individuals with these characteristics to be more capable of working (all else equal), this difference in characteristics will therefore reinforce Bound's expectation based on better health status, that the average labor supply of rejected applicants will be an upper bound for that of beneficiaries.

While the differences in education and age may reinforce the validity of the estimation strategy that we pursue above, it is important to note, however, that beneficiaries and rejected applicants may differ in other, unobserved characteristics. For example, at a given age and education level, those of above average health who decide to apply for benefits are more likely to have a lower cost or larger benefits of doing so. While in terms of many characteristics, rejected applicants may be more likely to work than beneficiaries this is not the case for all characteristics. 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.[6]

Finally, strictly speaking Bound's comparison method will identify an upper limit on the average effect of disability benefit receipt on the labor supply of applicants only. We will return to this issue later on.

The statistics presented in Table 2 indicate that in both our long and short run samples, not more than 27% of rejected applicants work. Those who work, work about 36 hours per week. Of beneficiaries fewer than 9% work in the survey month, and those who do, work about 33 hours per week. Together they imply that disability benefit receipt will reduce the labor force participation rates of applicants by at most 18 and 15 percentage points for the short and long run samples, respectively. The corresponding standard errors of these estimates are 1.8 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 17 (2.9) and 18 (1.6) percentage points, respectively. In terms of monthly hours of work, the estimated average reductions (and standard errors) associated with benefit receipt are 27 (3.1) and 23 (1.7) hours for the short and long run samples, and 26 (4.7) and 28 (2.6) for the short and long run vocational samples.

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 labor force participation rate of applicants. Our estimates of this effect are somewhat lower. There are several potential explanations for the difference in results. First, Bound's

[6]It is important to point out that while differences in unobservables may be a potential problem with the comparison group approach, this is not the case when using the second evaluation strategy pursued below, the RD approach.

sample only included older male applicants aged 45-64, while our samples include both male and female applicants aged 35 to 64. Second, our sample also includes applicants to the SSI program, a group that constitutes 36% 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.

Third, while Bound's data was collected in the 1970s, our data was 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.[7] 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 to lower the unemployment rate in the US by 0.5% (and by 4.2% for high school dropouts). The resulting compositional change in the pool of applicants to the DI program, with 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.

As mentioned earlier, a criticism of 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 fewer rejected applicants not working for these reasons. Moreover, as the time between the award and survey dates increases, the proportion of applicants who continue to be non-employed while continuing with appeals, should decline with subsequent denials. 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.

On the other hand, because our administrative data only cover award decisions made at the State level, some of the rejected applicants in our data may in fact have made successful appeals beyond the State level, and may therefore be misclassified beneficiaries. Of those denied benefits at the State level in 1995 29% went on to appeal beyond the State level, of which 63% were successful. In 1992 the corresponding percentages were 27% and 71% respectively (Green Book, 1993, 1996). This would imply that between 18% and 20% of

[7]For 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.

the rejected applicants in our data can be expected to have become beneficiaries through subsequent appeals. This misclassification leads to a downward bias in our estimates, with corrected estimates being approximately 25% larger than those reported earlier. This would lead to estimated labor force participation effects of between -19 and -23 percentage points for the various subsamples considered, estimates which are closer but still smaller (in absolute value) than Bound's estimates.

# 6 Evaluation Using A Regression Discontinuity Approach

As discussed earlier, 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 Figure 2. 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 Figure 2 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. For a full outline of the grid please see Figure 3.

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 they make 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 reliability of the RD approach we will adopt.

In order 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 around 55 years of age. 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.[8] 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 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 cutoff values found in the

[8]Note that the grid also contains discontinuities with respect to the number of years of education. 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.

grid. 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 of 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 left and right of the cutoff point.

The relationship between DI awards and age is explored in Figures 4 to 7. 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 SIPP members who applied sometime between 1989 and 2000. Even though the displayed award rate was smoothed, the main features of these graphs compare closely to those showing award rates by yearly and 3-months age intervals.[9] The vertical lines in each of these figures represent the Social Security grid cutoff ages of 45 (540 months), 50 (600 months) and 55 (660 months).

Figure 4 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 a vocational factor. While the graph reveals increases in the award rate around ages 50 and 55, they are relatively small. In our sample 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 5, 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 who face 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. Estimates of a piecewise linear probability model

[9]For confidentiality reasons, given the sample sizes involved, we were not allowed to report award rates by separate age levels.

specification of the award decision for the applications in the Extended sample reported in Table 3 mirror these patterns. The estimated jumps (and standard errors) in the award rate at the three cutoff ages were -0.012 (0.040) at age 45, 0.088 (0.045) at age 50 and 0.253 (0.037) at age 55.

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 the causal effect of treatment receipt $t_i$ (benefit receipt) on outcome $y_i$ (labor supply). Starting with the same notation from above, we can write:

$$y_i = \beta + \alpha_i t_i + u_i. \tag{2}$$

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

Consider first the simple case where treatment is a deterministic function of a continuous variable $A$, for example the individual's age, as in $t_i = 1\{A_i \geq \overline{A}\}$. The variable $A$ is likely to be directly correlated with the outcome variable $y$, thereby confounding the effect of $t_i$ on $y$ with that of $A$. This case is commonly referred to in the literature as the *Sharp Regression Discontinuity Design* (Thistlethwaite and Campbell, 1960).

The idea of comparing average outcomes just to the right and the left of the cutoff value $\overline{A}$, 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

$$\lim_{A\downarrow\bar{A}} E[y|A] - \lim_{A\uparrow\bar{A}} E[y|A] = E[\alpha_i|A = \bar{A}]. \tag{3}$$

The RD approach therefore identifies a local average treatment effect $E[\alpha_i|A = \bar{A}]$.[10] 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 not receiving disability benefits.

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 probability function that will be discontinuous in age at the cutoff values found in the grid. In the fuzzy RD case the propensity score function $E[t_i|A]$ is discontinuous in $A$ at a known value $\overline{A}$. As before this design allows us to identify

[10]Strictly speaking, this result only requires right-continuity in assumption S2.

an average treatment effect. To see this, consider the case where the treatment effect is a constant ($\alpha_i = \alpha$), then at the age cutoff $\overline{A}$

$$(4) \qquad \begin{aligned} \lim_{A\downarrow\overline{A}} E[y_i|A] - \lim_{A\uparrow\overline{A}} E[y_i|A] \;&=\; \alpha \cdot \Big[\lim_{A\downarrow\overline{A}} E[t_i|A] - \lim_{A\uparrow\overline{A}} E[t_i|A]\Big] \\ &\quad + \lim_{A\downarrow\overline{A}} E[u_i|A] - \lim_{A\uparrow\overline{A}} E[u_i|A]] \end{aligned}$$

It follows that under assumptions S1 and S2 the treatment effect $\alpha$ is identified by

$$(5) \qquad \frac{\Big(\lim_{A\downarrow\overline{A}} E[y_i|A] - \lim_{A\uparrow\overline{A}} E[y_i|A]\Big)}{\Big(\lim_{A\downarrow\overline{A}} E[t_i|A] - \lim_{A\uparrow\overline{A}} E[t_i|A]\Big)}.$$

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

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 $\alpha_i$ conditional on $A$ near $\bar{A}$, the ratio above identifies the local average treatment effect $E[\alpha_i|A = \bar{A}]$. 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:

$$(6) \qquad \frac{\lim_{A\downarrow\bar{A}} E[y|A] - \lim_{A\uparrow\bar{A}} E[y|A]}{\lim_{A\downarrow\bar{A}} E[t|A] - \lim_{A\uparrow\bar{A}} E[t|A]} = \lim_{e\downarrow 0} E[\alpha_i | t_i(\bar{A}+e) - t_i(\bar{A}-e) = 1]$$

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.

Before discussing the estimation of these average treatment effects, it is important to consider the validity of the underlying continuity conditions in our application. The assumptions that $E[u_i|A]$ 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). If 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 design. However, given that the rule, while commonly used by DI administrators in evaluating disability cases, is only used for internal purposes and not advertized, 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 became eligible for early retirement benefits at age 55, 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. In both cases we would expect to see a jump in the application rate at the cutoff ages. Figure 8 which plots the percentage of applicants in our sample who applied at each given age shows no evidence of heaping near the cutoff ages.

To estimate the average treatment effect defined above, we adopt two different estimation methods: a semiparametric two-stage method proposed by Van der Klaauw (2002) and the nonparametric 'local Wald' estimator proposed by Hahn et al (2001). In case of the latter, each of the limits in (6) at each cutoff age $\overline{A}_j$ is estimated using a one-sided uniform kernel estimator. More specifically, the limits are estimated as,

$$\lim_{A\downarrow\overline{A}_j} E[y_i|A] = \frac{\sum_{i\in S} y_i w_i}{\sum_{i\in S} w_i} \qquad \lim_{A\uparrow\overline{A}_j} E[y_i|A] = \frac{\sum_{i\in S} y_i(1-w_i)}{\sum_{i\in S}(1-w_i)}$$
$$\lim_{A\downarrow\overline{A}_j} E[t_i|A] = \frac{\sum_{i\in S} t_i w_i}{\sum_{i\in S} w_i} \qquad \lim_{A\uparrow\overline{A}_j} E[t_i|A] = \frac{\sum_{i\in S} t_i(1-w_i)}{\sum_{i\in S}(1-w_i)},$$

where $w_i$ is an indicator for whether the observation is above or equal to the age cutoff such that $w_i = I(\overline{A}_j \leq A_i < \overline{A}_j + h)$ where $h$ is a bandwidth, and $S$ denotes the subsample around the cutoff point defined by $(\overline{A}_j - h \leq A_i < \overline{A}_j + h)$. This nonparametric estimator is therefore numerically equivalent to a Wald estimator applied to the subsample S, where $t_i$ is instrumented by $w_i$.[11].

The two-stage estimation procedure proposed by Van der Klaauw (2002) involves the estimation of a control function augmented labor force participation equation in which the treatment variable is replaced by an estimated propensity score. More formally, in the first stage, the propensity score is estimated as,

$$E[t_i|A_i] = 1|A_i) = g(A_i) + \sum_{j=1}^{3} \gamma_j \cdot 1\{A_i \geq \bar{A}_j\} \tag{7}$$

where $g(A_i)$ is a flexible continuous function in $A_i$, and the $\gamma_j$ represent the discontinuities in the award rate due to the rules of the DI program.

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

[11]However, 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).

association between age and labor force participation. The second stage equation is given by

$$y_i = \beta + \delta t_i + k(A_i) + w_i, \tag{8}$$

where in estimation $t_i$ is replaced by the first stage estimate of $E[t_i|A_i]$. 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), where the order of the series is increased until the fit no longer improves.

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 a function of $A$ over their whole domains, it also assumes them to be differentiable over its domain.[12] As discussed in Van der Klaauw (2002), the estimate of $\delta$ represents an estimate of a (weighted) average of the three average local treatment effects $E[\alpha_i|\overline{A}_j]$.

The two-stage estimates reported in Table 4 are based on a first stage piecewise linear specification of the award equation.[13] The table shows estimates for linear and quadratic specifications of the control function $k(A)$, as well as for the series approximation. While in the case of labor force participation the optimal order of the series was determined to be 2 (a quadratic specification), for the monthly hours of work outcome it was one, both in case of the long run and short-run samples.

The first two RD estimates presented in the third column of Table 4, for the short-run and long-run samples, respectively, are comparable to the upper bound estimates of -0.17 and -0.18 we presented earlier. While not statistically significant, the standard errors are small enough to exclude, using a 95% confidence level, effects bigger (in absolute terms) than -0.38. Somewhat surprisingly, we find the estimates based on the long-run sample to be somewhat smaller in absolute size than those for the short-run sample, although this difference is not statistically significant. 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 above the State level (and would return to work after final dismissal of their applications) 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

[12]This approach is similar to the recent Robinson-type estimator proposed by Porter (2003) which also assumes continuity and differentiability of $E[u|A]$.

[13]Estimates for a piecewise quadratic specification yielded similar results.

low costs of applying, and with a low commitment to the labor force.

Also presented in the third and fourth rows of Table 4 are LFP effect estimates where $t_i$ in equation (5) was instrumented using an estimate of $E[t_i|A_i]$ obtained from the same piecewise linear model presented above but now estimated on the much larger 'Extended' sample of all Stage 5 applications contained in our 831 file (estimates of which are presented in Table A1).[14] The estimates and standard errors are very similar to those in the first two rows.

The second part of Table 4 presents the estimated impact of benefit receipt on monthly hours of work. While the LFP estimates shown in the first part of the table were slightly smaller in absolute value than the upper bound estimates, the RD estimates for monthly hours of work are instead a little larger, with benefit receipt estimated to reduce hours of work by between 33 and 51 hours per month.[15] The estimates based on the short run sample are again a little larger in absolute value than those for the long-run sample.[16] Combined the estimates imply that DI benefit receipt leads to a modest, but statistically significant reduction in the percentage of people who work and in average monthly hours of employment.

Table 5 reports local Wald estimates for the third cutoff at age 55 and presents DI impact estimates for labor force participation and hours of work. With the exception of the short-run sample estimates that correspond to a regression-discontinuity sample based on a two year age interval (which are based on small sample sizes and have large standard errors), the local Wald estimates of the LFP effects are comparable to the RD estimates presented in Table 4. The estimates for hours of work, on the other hand, are somewhat smaller and instead more similar to the upper bound estimates presented earlier.

As shown in Table 5, if we compare individuals within 4 points below and above the 55 year old cutoff point for the long run sample, it is estimated that DI benefit receipt decreases labor supply by 20 hours per month and participation by 12 percentage points. If we decrease the bandwidth to smaller intervals i.e. 3 years or 2 years above and below the cutoff the results are very similar. The results for the short run sample are a lot more sensitive to the specification of bandwidth but this a reflection of the small sample sizes used in the estimation.

---

[14] More specifically, we estimated the piecewise linear award equation using the sample of all Stage 5 applications in the 831 file. These estimates were used to calculate, for each applicant in our matched samples, the predicted probability of receiving an award, which was subsequently used to instrument $t_i$ in equation (5). We are grateful to Josh Angrist for suggesting this estimation approach.

[15] Equivalently, the estimates imply that without benefit receipt applicants would on average have worked an extra 33 to 51 hours per month.

[16] In a series of sensitivity tests we found these estimates to be insensitive to the specification of the award equation.

We also compared individuals at the other two age cutoffs of 45 and 50. These results are not presented because the numbers of observations are too small to obtain precise enough estimates for either cutoff point. Similarly, we obtained estimates of a pooled Wald estimator that is a weighted average of the three local Wald estimators and found these to be very close to the estimates solely based on the third cutoff.

# 7 Conclusion

We presented estimates of 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 applicants would have been at most 23 percentage points higher had none received benefits compared to the case where all received benefits. They also imply that the increase in the labor force participation rate among beneficiaries had they not received benefits would have been at most 23 percentage points.

We also apply a regression discontinuity approach to find even smaller labor supply responses for a group of 'marginal' applicants whose disability determination is based on vocational factors. The RD estimates suggest that DI decreases labor supply by 20 hours per month and participation by 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% (1993 Greenbook) to 37% (Lahiri et al, 1995), and in our sample it represents 40% of all applicants. 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. Increasing these 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.

Together these findings suggest that during the 1990s the work disincentive effects of the DI program were rather modest: the vast majority of applicants would not have worked even if none had received DI. These estimates are slightly lower than those found by Bound for the 1970s. we speculate the smaller effects are a reflection of the differences in the time periods under study. Two of these are outlined below.

First during the rules of the DI program changed. In the earlier part of the nineties, the requirements to qualify for DI became much more lenient than in previous decades. This leniency coupled with poor labor market conditions in the early and middle of the decade, particularly for low-income workers, may have changed the composition of people who apply for DI. The decreased opportunities for low-skilled workers in the labor force coupled with a decreasing real wage and increasing real disability benefit amounts for this segment of the labor force may have induced workers to apply for disability (Juhn et al 2002). If these applicants have lower labor force attachment then past cohorts then this could account for the smaller impact estimates that we find. Whether they would not work because of their medical conditions, or because of unfavorable labor market conditions, is beyond the scope of this paper, but is an important area for further inquiry.

Second, in the period between 1970 (when Bound's data was collected) and the nineties, there has been significant increase in the labor supply of married women. If the presence of these women in the home creates a source of income maintenance than this would be another explanation for the lower estimates compared to Bound. We are currently exploring the role that spouses play. Preliminary findings suggest that find that DI benefits affect spousal labor supply differentially. For males, DI receipt decreases spousal labor supply by 30 percentage points. The effect is ten percentage points lower for females. Preliminary investigation into the effect of DI on joint labor force participation suggests that DI receipt decreases couple's labor force participation by 20 percentage points.

In terms of spending, the DI 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. The RD estimates provide an idea of the impact that increasing the age cutoffs would have on labor force participation. Increasing these 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.

It is important for policymakers to understand the impact of welfare programs on both the individual and the household. This research represents one of the first forays into this arena.

## Bibliography


Autor, David H. and Mark G. Duggan, “The Rise in the Disability Rolls and the Decline in Unemployment”, *Quarterly Journal of Economics*, 2003, 157-206.

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

Bound, John, “The Health and Earnings of Rejected Disability Insurance Applicants: Reply”, *American Economic Review* 81(5), December 1991, 1427-1434.

Bound, John and Richard V. Burkhauser, “Economic analysis of transfer programs targeted on people with disabilities” in *Handbook of labor economics*, Volume 3A,eds. Ashenfelter, Orley and David Card, Vol. 3C. Amsterdam; New York and Oxford: Elsevier Science, 2000.

Bound, John, Richard V. Burkhauser, and A. Nichols, “Tracking the Household Income of SSDI and SSI Applicants”, Population Studies Center Research Report, University of Michigan, No. 02-502, 2002.

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

Burtless, Gary, “An Economic View of Retirement ” in *Behavioural Dimensions of Retirement Economics*, ed Aaron, Henry J., Brookings Institution Press Washington DC and Russell Sage Foundation, New York, 1999.

Chirikos, Thomas N, “The Relationship Between Health and Labor Market Status”, *Annual Review of Public Health* 14, 1993, 292-312.

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

Hahn, Jinyong, Petra Todd and Wilbert van der Klaauw, “Identification and Estimation of Treatment Effects with a Regression- Discontinuity Design”, *Econometrica*, 69(1), 2001, pages 201-209.

Haveman, Robert H. and Barbara L. Wolfe, “The Decline in Male Labor Force Participation: Comment”, *Journal of Political Economy*, 92(3), June 1984a, 532-541.

Haveman, Robert H. and Barbara L. Wolfe, “Disability Transfers and Early Retirement: A Causal Relationship?” *Journal of Public Economics*, 24, 1984b, 47-66.

Hoynes, Hilary W. and Robert Moffitt, “Tax Rates and Work Incentives in the Social Security Disability Insurance Program: Current law and Alternative Reforms”, *National Tax Journal*, December 1999.

Hu, Jianting, K. Lahiri, D.R. Vaughan, and B. Wixon “A Structural Model of Social Security’s Disability Determination Process”, *The Review of Economics and Statistics*, 83 (2), 2001, 348-361.

Hu, Jianting, K. Lahiri, D.R. Vaughan, and B. Wixon, "A Structural Model of Social Security's Disability Determination Process ", *Social Security Administration Working Paper*, 72, 1997.

Juhn, Chinhui, "Current Employment, Historically Contemplated", *Brookings Papers on Economic Activity*, 1, 2002

Lambrinos, James, "Health: A Source of Bias in Labor Supply Models", *The Review of Economics and Statistics*, 63 (1), 1981, 206-212.

Lahiri, Kajal, Denton R. Vaughan, and Bernard Wixon, "Modelling SSA's Sequential Disability Determination Process Using Matched SIPP Data", *Social Security Bulletin* Vol. 58, Winter 1995.

Leonard, Johnathan, "Labor Supply Incentives and Disincentives for Disabled Persons" In *Disability and the Labor Market: Economic Problems, Policies and Programs*, eds. Monroe Berkowitz and M. Anne Hill, Utica, NY: Industrial and Labor Relations Press, 1986.

Newey, W. K., J.L. Powell, and J. R. Walker, "Semiparametric Estimation of Selection Models: Some Empirical Results", *American Economic Review* 80, 1990, 324- 28.

Olsen, Janice A., "Social Security Benefit Reporting In The Survey Of Income And Program Participation And In Social Security Administration Records", Unpublished Manuscript, Division of Economic Research, Office of Research, Evaluation, and Statistics, Office of Policy, Social Security Agency, June 27, 2001.

Parsons, Donald O., "Health, Family Structure, and Labor Supply", *The American Economic Review*, 67(4), September 1977, 703-712.

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

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

Porter, Jack, "Estimation in the Regression Discontinuity Model" manuscript, Harvard University, 2003.

Rupp, K. and D. C. Stapleton, *Growth in Disability Benefits. Explanations and Policy Implications*, eds. Social Security Administration and The Lewin Group, 1998.

Slade, F. P., "Older Men: Disability Insurance and the Incentive to Work", *Industrial Relations* 23(2), Spring 1984, 260-69.

Stern, S., "Measuring the Effect of Disability on Labor Force Participation", *Journal of Human Resources*, 24(3), 361-395.

Thistlethwaite, D. and D. Campbell, "Regression- Discontinuity Analysis: An alternative to the ex post facto experiment", *Journal of Educational Psychology*, 51, 1960, 309-317.

U.S. House of Representatives, committee on Ways and Means (HRWCWM), *1993 Green-*

*book: Overview of Entitlement Programs*, U.S. Government Printing Office Washington D.C., 1993.

U.S. House of Representatives, committee on Ways and Means (HRWCWM), *1996 Greenbook: Overview of Entitlement Programs*, U.S. Government Printing Office Washington D.C., 1996.

U.S. House of Representatives, committee on Ways and Means (HRWCWM), *Greenbook: Overview of Entitlement Programs*, U.S. Government Printing Office, Washington D.C., 2000.

U.S. Social Security Administration, *Social Security Bulletin Annual Statistical Supplement*, Various Years.

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

Table 1: Sample Statistics

| Variable | All Applicants | | Stage 5 Applicants | |
|---|---|---|---|---|
| | Short Run | Long Run | Short Run | Long Run |
| Age (months) | 604.5 | 614.1 | 605.0 | 617.6 |
| High School Diploma | 0.66 | 0.61 | 0.67 | 0.59 |
| Labor Force Partipation | 0.20 | 0.16 | 0.17 | 0.14 |
| Monthly Hours Worked | 27.7 | 23.0 | 24.0 | 20.3 |
| Months Since Award Date | 22 | 47 | 22 | 48 |
| Number Observations | 1583 | 4603 | 640 | 1807 |

Table 2: Characteristics by Award Status

| | All Applicants | | | | Stage 5 Applicants | | | |
|---|---|---|---|---|---|---|---|---|
| | Short Run | | Long Run | | Short Run | | Long Run | |
| | $T = 0$ | $T = 1$ | $T = 0$ | $T = 1$ | $T = 0$ | $T = 1$ | $T = 0$ | $T = 1$ |
| Age (months) | 582.6 | 631.5 | 593.4 | 635.0 | 556.8 | 654.4 | 559.8 | 664.1 |
| HS Diploma | 0.67 | 0.65 | 0.61 | 0.60 | 0.72 | 0.62 | 0.64 | 0.55 |
| LFP | 0.27 | 0.09 | 0.23 | 0.08 | 0.26 | 0.09 | 0.24 | 0.06 |
| Monthly Hours | 38.5 | 11.8 | 34.6 | 11.3 | 36.9 | 10.8 | 36.0 | 7.6 |
| Number Obs | 875 | 708 | 2305 | 2298 | 324 | 316 | 805 | 1002 |

Beneficiaries are indicated by $T = 1$ while rejected applicants are denoted by $T = 0$.

Table 3: Estimates of Award Equation - Extended Sample

| Coefficient | Estimate | Standard Error |
|---|---|---|
| $\gamma_1$ | -0.013 | (0.040) |
| $\gamma_2$ | 0.088 | (0.045) |
| $\gamma_3$ | 0.253 | (0.037) |
| $\psi_{00}$ | 0.414 | (0.163) |
| $\psi_{01}$ | -0.003 | (0.003) |
| $\psi_{11}$ | 0.092 | (0.099) |
| $\psi_{12}$ | 0.247 | (0.131) |
| $\psi_{13}$ | -0.208 | (0.097) |
| Obs | 4343 | |

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

Table 4: RD Estimates of Labor Supply Effects of DI Receipt

| Sample | First Stage | Linear | Quadratic | Series Approximation |
|---|---|---|---|---|
| | | Labor Force Participation | | |
| SR | Matched | -0.129<br>(0.122) | -0.269<br>(0.183) | -0.129<br>(0.122) |
| LR | Matched | -0.092<br>(0.068) | -0.191<br>(0.089) | -0.092<br>(0.068) |
| SR | Extended | -0.179<br>(0.122) | -0.323<br>(0.167) | -0.179<br>(0.122) |
| LR | Extended | -0.090<br>(0.068) | -0.190<br>(0.090) | -0.090<br>(0.068) |
| | | Hours of Work | | |
| SR | Matched | -19.91<br>(19.53) | -45.57<br>(29.32) | -45.57<br>(29.32) |
| LR | Matched | -16.35<br>(10.89) | -33.15<br>(14.31) | -33.15<br>(14.31) |
| SR | Extended | -26.54<br>(19.44) | -50.82<br>(26.76) | -50.82<br>(26.76) |
| LR | Extended | -16.04<br>(10.88) | -32.99<br>(14.37) | -32.99<br>(14.37) |

Heteroskedasticity consistent standard errors in parentheses. They have been corrected for generated regressors.

Table 5: Local Wald Estimates of Labor Supply Effects of DI Receipt

| Sample | 2 year interval | 3 year interval | 4 year interval |
|---|---|---|---|
| | Labor Force Participation | | |
| SR | 0.049 | -0.089 | -0.141 |
| | (0.142) | (0.104) | (0.102) |
| Obs | 102 | 136 | 187 |
| LR | -0.146 | -0.134 | -0.118 |
| | (0.105) | (0.069) | (0.062) |
| Obs | 289 | 411 | 552 |
| | Hours of Work | | |
| SR | -2.492 | -11.620 | -22.018 |
| | (21.839) | (14.656) | (14.713) |
| Obs | 102 | 136 | 187 |
| LR | -22.898 | -22.568 | -20.381 |
| | (15.666) | (10.545) | (8.836) |
| Obs | 289 | 411 | 552 |

# Appendix of Figures

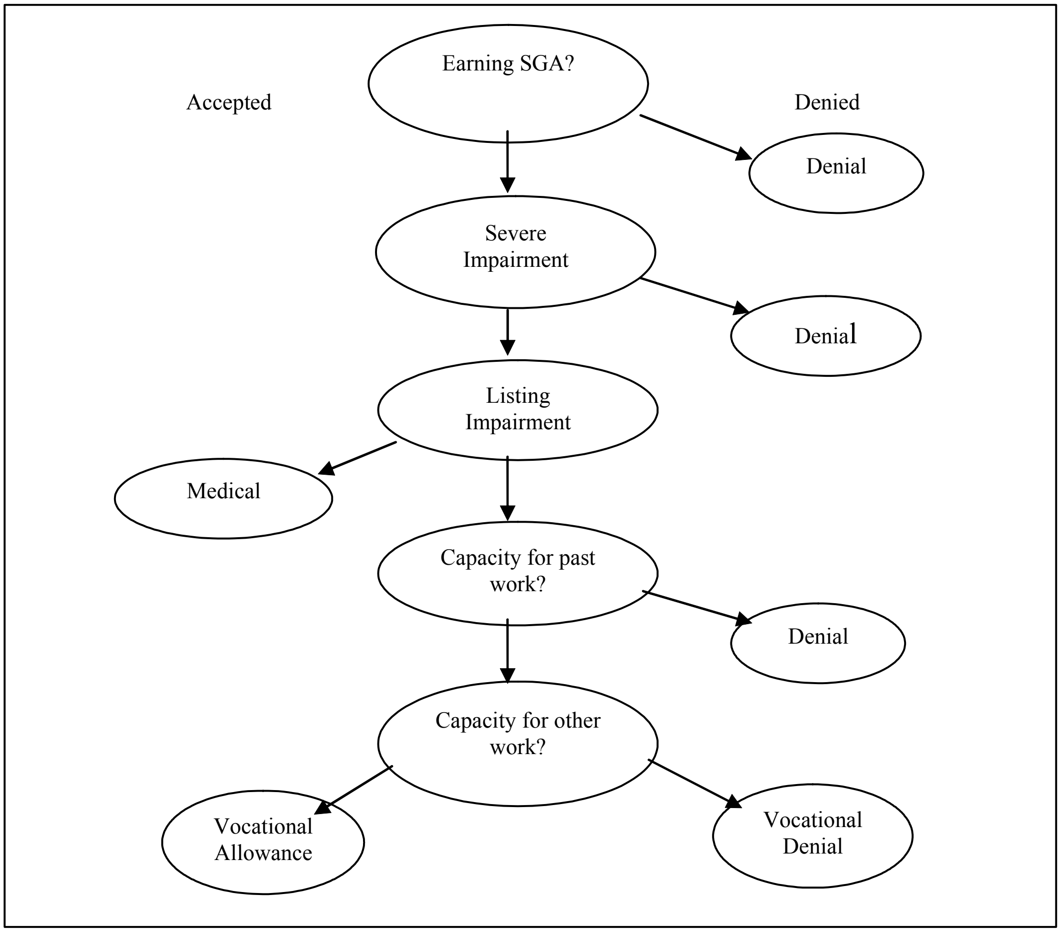


**Figure 1: The 5 Stages of the Social Security Disability Determination Process**

| 7th Through 11th Grade | | | | | |
|---|---|---|---|---|---|
| Residual Functional Capacity | Age | Unskilled/None | | (Semi) Skilled/Non-transferable | |
| | 60-64 | D | | D | |
| | 55-59 | D | | D | |
| SEDENTARY | 50-54 | D | * | D | * |
| | 45-49 | N | | N | |
| | 18-45 | N | | N | |
| | | | | | |
| | 60-64 | D | | D | |
| | 55-59 | D | * | D | * |
| LIGHT | 50-54 | N | | N | |
| | 45-49 | N | | N | |
| | 18-45 | N | | N | |
| | | | | | |
| | 60-64 | D | | N | |
| | 55-59 | N(D) | | N | |
| MEDIUM | 50-54 | N | | N | |
| | 45-49 | N | | N | |
| | 18-45 | N | | N | |

**Figure 2: Excerpt from the Medical Vocational Grid**
**(D=Disabled, N=Not Disabled)**

| 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 | |

**Figure 3: The Medical Vocational Grid (D=Disabled, N=Not Disabled)**

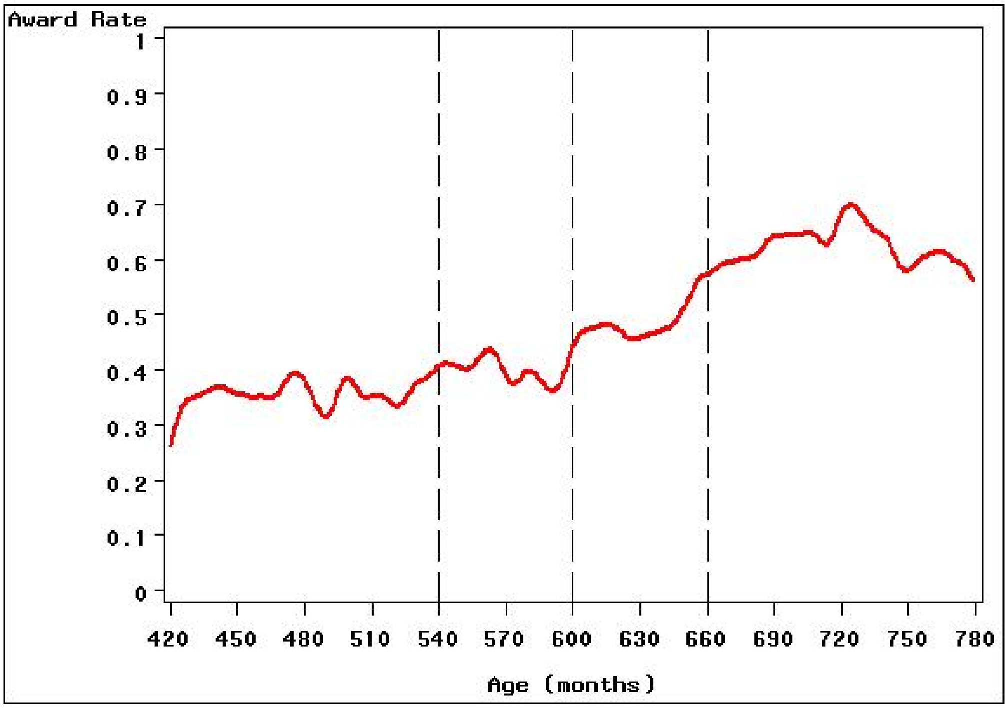


**Figure 4: Disability Insurance Award Rates – All Applicants**

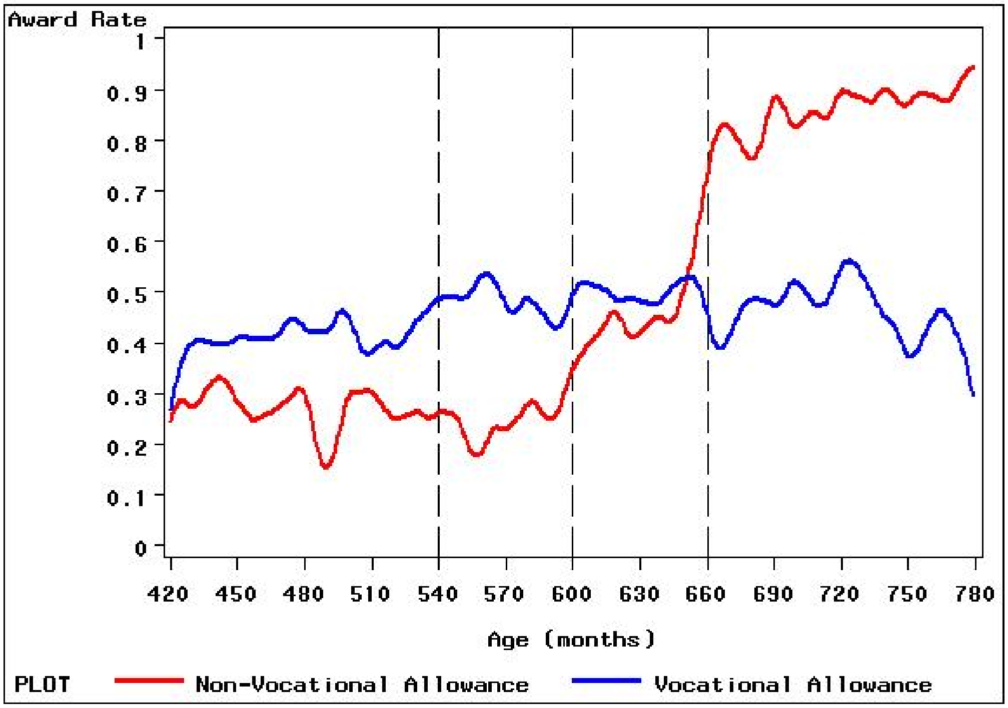


**Figure 5: Award Rate By Determination Stage – All Applicants**

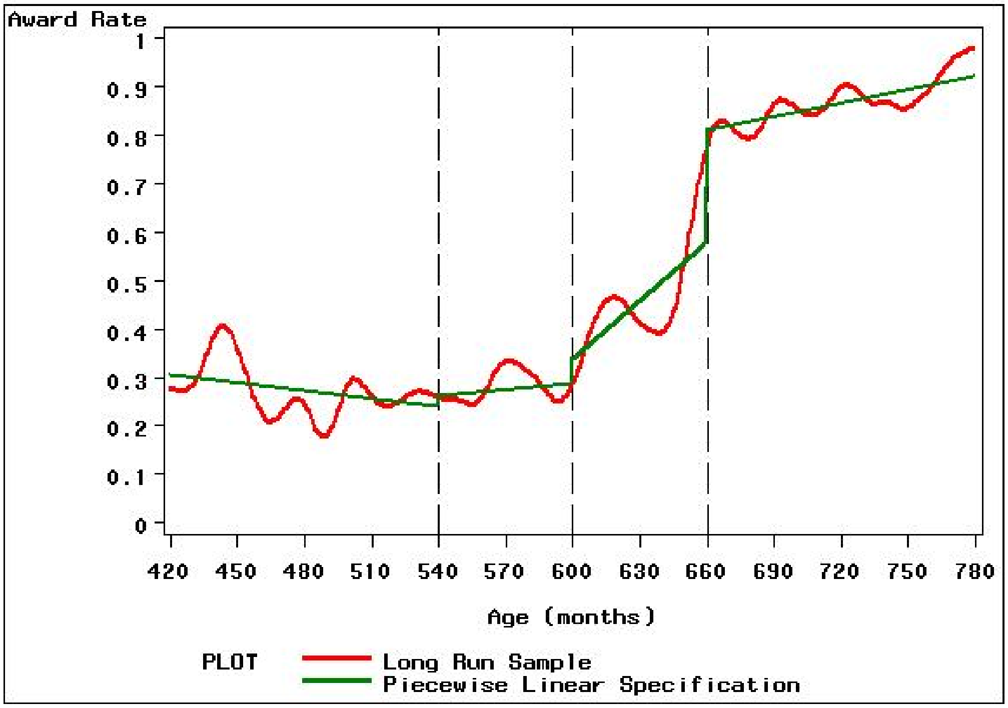


**Figure 6 : Award Rates for Stage 5 Applicants – Long Run Sample**

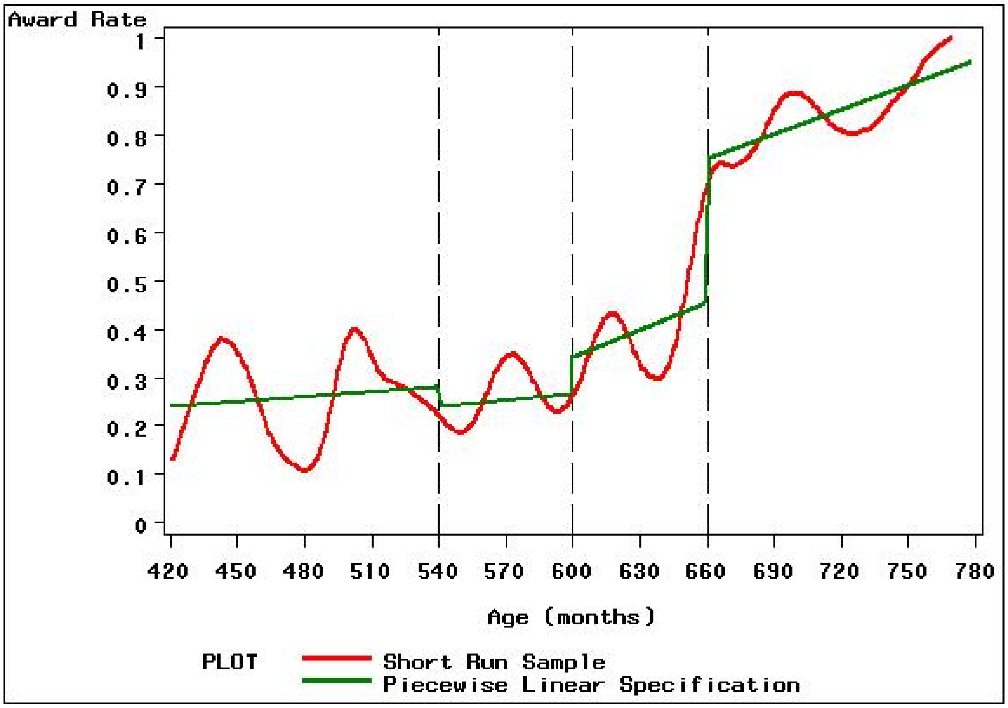


**Figure 7 : Award Rates For Stage 5 Applicants - Short Run Sample**

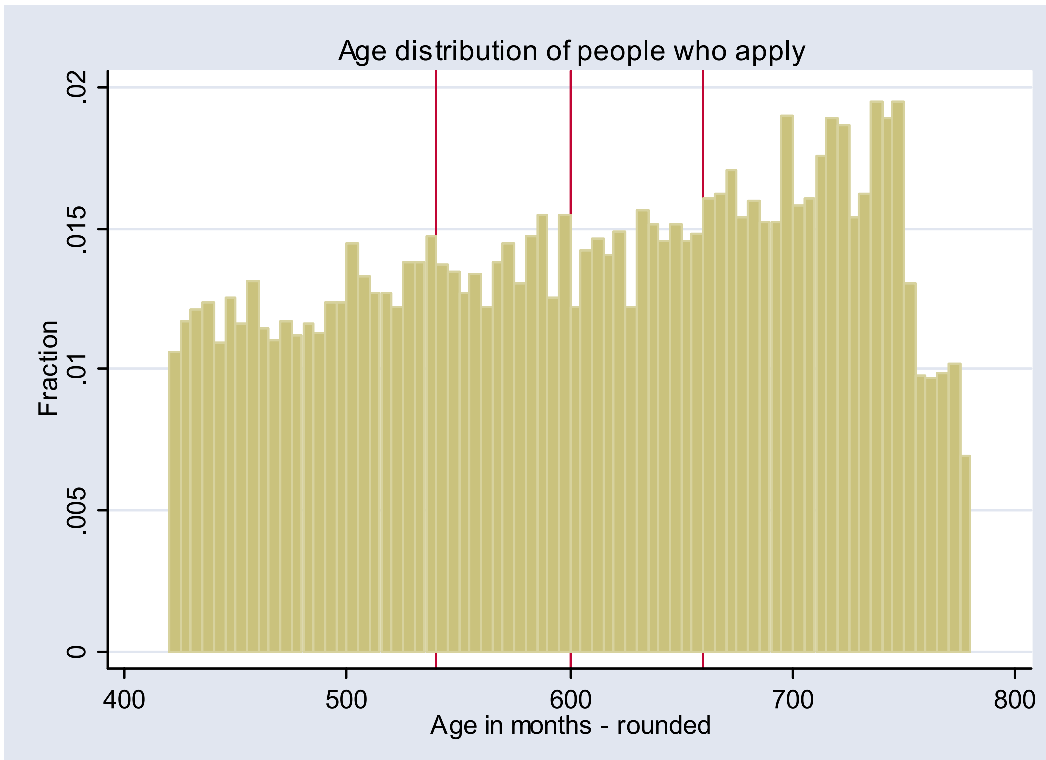


**Figure 8: Age Distribution of All Applicants to DI Program**