# Economic Consequences of Tightening the Disability Screening*

David Autor† Andreas Ravndal Kostøl‡ Magne Mogstad§

September 2014

**Abstract:** A common policy proposal to reverse the growth in disability insurance (DI) rolls is to tighten the screening process. On the one hand, stricter screening criteria could reduce the number of individuals who still have substantial work capacity from collecting benefits. On the other hand, it could result in income or welfare loss to individuals and families who value the insurance aspect of the program in cases of true disability. To date, most existing research focus on the labor market impact of DI receipt, and little is known about how a tightening of the disability screening would affect individuals and families on other margins. This paper helps to close that gap by analyzing the broader economic consequences of tightening the disability screening. The key to our research design is twofold. First, we are able to link a number of administrative registers from Norway, providing a unique population panel data set with detailed information about every individual and family. Second, the DI system in Norway randomly assigns judges to DI applicants whose cases are initially denied; some appeal judges are systematically more lenient, which leads to random variation in the probability an individual will get onto DI. We use the random assignment of appeal judges to estimate the causal effects of DI allowance on a variety of outcomes, including not only individual labor force participation and earnings but also participation in other social insurance programs and the income available for household consumption.

**Keywords:** disability insurance; labor supply; benefit substitution, disposable income
**JEL codes:** I38, J62, H53

*This research was supported by the U.S. Social Security Administration through grant #1 DRC12000002-02-00 to the National Bureau of Economic Research as part of the SSA Disability Research Consortium. The findings and conclusions expressed are solely those of the author(s) and do not represent the views of SSA, any agency of the Federal Government, or the NBER. We are grateful to Knut Brofoss, Espen Vihle and Runar Narvland for their help in accessing the data and in understanding the institutional details.

†MIT Department of Economics; NBER. E-mail: dautor@mit.edu

‡Research Department, Statistics Norway; Department of Economics, University of Bergen. E-mail: kos@ssb.no

§Department of Economics, University of Chicago; Research Department, Statistics Norway; NBER. E-mail: magne.mogstad@gmail.com

# 1 Introduction

Over the past 50 years, disability insurance (DI) rolls have risen dramatically in many OECD countries. In the U.S., the participation rate has increased from less than 1% to over 5% of the adult population. In many European countries, the increases are even more striking, from 1% to 7% in the U.K and from 2% to almost 10% in Norway. These increases have made DI one of the largest transfer programs in most OECD countries. In the U.S., for example, outlays for DI exceed those for food stamps, traditional cash welfare, or the EITC.[1] For families without small children, DI is often the only cash benefit available after unemployment benefits run out and it has therefore become an increasingly important component of the social safety net.

A key reason for the growth in DI rolls is the liberalization of the screening process (Autor & Duggan, 2003, 2006). The liberalization made DI more accessible to individuals with difficult-to-very disorders, by placing less weight on diagnostic and medical factors and relatively more weight on the ability to function in a work setting. This shift towards subjective factors in disability determination decisions has fueled a debate over whether the DI program today is primarily being used as a gateway for early retirement, rather than providing insurance against health shocks that prevent work. While some countries have already tightened the eligibility criteria to reverse the growth in DI rolls, other countries are debating similar policy changes.[2]

Two key questions in thinking about policy proposals to tighten the screening process are to what extent the DI program discourages work, and how valuable the insurance is to individuals and families. On the one hand, stricter screening criteria could reduce the number of individuals who still have substantial work capacity from collecting benefits. On the other hand, it could result in income or welfare loss to individuals and families who value the insurance aspect of the program in cases of true disability.

Assessing this insurance-incentive tradeoff in tightening the disability screening has proven difficult for several reasons. It is often difficult to access data on disposable income or consumption which can be linked with social security records. Another key challenge is that individuals receiving DI are likely to differ from non-recipients in important ways, both in observable and unobservable dimensions. To date, most empirical research focus on the labor market impact of DI receipt, and less is known about how a tightening of the disability screening would affect individuals and families on other margins.

This paper helps to close that gap by analyzing the broader economic consequences of tightening the disability screening. The key to our research design is twofold. First, we are able to link a number of administrative registers from Norway, providing a unique population panel data set with detailed

[1] In 2011 the U.S. paid out $129 billion to 10.6 million disabled workers and their families, with an additional $33 billion worth of disability benefits from the SSI program for poor Americans and $90 billion in Medicaid for disabled workers (OASDI Trustees Report, 2012). In 2009, DI payments constituted 1.8% of GDP in the U.S. and 2.3% of GDP across the European OECD-countries (OECD, 2010).

[2] For example, the U.S. tightened the criteria for new disability awards in the late 1970s and introduced an aggressive program of continuing disability reviews in 1980; however, Congress responded by halting the reviews and, in 1984, liberalizing the program's screening criteria along several dimensions. Another example is the Netherlands; in 1994, the eligibility criteria were tightened and the growth in DI rolls reversed.

information about every individual and family. Second, the DI system in Norway randomly assigns judges to DI applicants whose cases are initially denied; some appeal judges are systematically more lenient, which leads to random variation in the probability an individual will get onto DI. We use the random assignment of appeal judges to estimate the consequences of tightening the disability screening.[3]

As our measure of judge leniency, we use the average allowance rate in all the other cases a judge has handled. This leniency measure is highly predictive of the judge's decision in the current case, but as we document, uncorrelated with observable case characteristics. Using this random variation as an instrument, we estimate the causal effects of DI allowance on a variety of outcomes, including not only individual labor force participation and earnings but also participation in other social insurance programs and the income available for household consumption. To assess the internal validity of our research design, we perform a number of robustness checks, all of which suggest the identifying assumptions of independence, exclusion and monotonicity hold. In ongoing work, we are collecting additional data on assets and consumption, which would allow us to estimate the degree of insurance provided by the DI program, over and above self-insurance through savings and the spouse's labor supply.

Our paper contributes to a growing literature on the causes and consequences of the growth in DI rolls (for a review, see Autor & Duggan, 2006; Autor, 2011a). To date, research has largely focused on estimating the work capacity and labor supply elasticity of DI recipients.[4] Yet despite a recent surge in research on this topic, less is known about how the receipt of DI affects individuals on margins other than labor force participation.[5] Our study provides some of the first causal evidence on the broader economic consequences of reducing DI participation rates by tightening the disability screening. Our estimates point to the importance of benefit substitution in attenuating the consequences of changes in disability screening, at least in the context of a fairly generous European welfare state. However, even in the presence of benefit substitution, we find that stricter disability screening causes a substantial drop in the disposable income of already poor families.

Two studies using U.S. data and a similar research design have looked at how DI receipt affects labor supply. Maestas *et al.* (2013) use variation in the leniency of initial examiners in the U.S. and find that DI receipt substantially reduces earnings and employment of applicants. Exploiting the

[3] A similar identification approach based on the quasi-random assignment of judges (or examiners) has been used in several contexts, such as to study the labor supply effects of DI receipt (Maestas *et al.* , 2013; French & Song, 2013), the impacts of incarceration (Kling, 2006; Aizer & Joseph J. Doyle, 2013), the consequences of foster care (Doyle (2007, 2008)), the effects of consumer bankruptcy protection (Dobbie & Song (2013)), and the intergenerational transmission of welfare receipt (Dahl *et al.* , 2013)

[4] See e.g. Autor *et al.* (2003); Borghans *et al.* (2012); Bound (1989); Campolieti & Riddell (2012); French & Song (2013); Gruber (2000); Kostol & Mogstad (2014); Maestas *et al.* (2013); Moore (2011); Parsons (1991); von Wachter *et al.* (2011).

[5] There are some notable exceptions. Borghans *et al.* (2012), De Jong *et al.* (2011), Karlstrom *et al.* (2008), and Staubli (2011) analyze how reforms of the DI system affect take-up rates of certain other programs. None of these studies look at the impacts of DI receipt on the income available for household consumption. Low & Pistaferri (2012) provide simulations from a calibrated a lifecycle model to compare the insurance value and incentive costs of DI benefits. Other studies document the decline in income and consumption that follow health changes or disability (see e.g. Meyer & Mok, 2013). Dahl *et al.* (2013) estimate how DI allowance in one generation causes DI receipt in the next generation.

leniency of appeal judges in the U.S., French & Song (2013) find comparable labor supply effects of DI receipt among appellants. What makes our study unique is the ability to link the judicial decisions to a wide range of variables for appellants and their family members. This allows us to provide novel evidence on the insurance value and incentive costs of DI benefits, in a setting where we can credibly address concerns about omitted variables bias.

At the same time, it is important to emphasize the local nature of our results. Our IV estimates are local average treatment effects for individuals who could have received a different allowance decision in the appeal process had their case been assigned to a different judge. Our instrument picks out these compliers who are on the margin of program entry. This means we need to be cautious in extrapolating the causal effects we estimate to the population at large or to other settings. For example, the work capacity of individuals who are at the margin of program entry are likely to differ from that of infra-marginal individuals.

At the same time, the economic consequences among the compliers to our instrument are relevant for policy, since reforms aimed at stemming the rise in DI will likely have the largest effect on applicants on the margin of program entry. In both Norway and the U.S., the rise in DI rolls in recent decades appears to be primarily driven by a more liberal screening of marginal applicants who are often initially denied and relatively likely to appeal (Autor & Duggan, 2006; Kostol & Mogstad, 2014). Our estimates serve to highlight that benefit substitution and spouse's labor supply response can be important margins to capture for accurate projections of how reforms of the DI program affect participation rates and program costs of the broader social security system.

The remainder of the paper proceeds as follows: Section 2 reviews the key features regarding the DI program in Norway, compares the system with the U.S. system, and describes the experimental research design. In Section 3, we describe the data and sample restrictions. Section 4 assesses the relevance and validity of our instrument, before Section 5 presents our empirical findings. The final section offers some concluding remarks.

## 2 Background

In this section, we begin by reviewing key facts regarding the DI program in Norway. We then provide empirical evidence on the disability determination process, documenting in particular that the system generates random variation in DI awards. We further describe how our empirical model uses this exogenous variation to estimate the economic consequences of DI allowance.

### The Norwegian DI program

In Norway, DI benefits are designed to provide partial earnings replacements to all workers under the full retirement age who are unable to engage in substantial gainful activity because of a medically determined physical or mental impairment that has lasted for at least a year. The DI program is part of the broader Social Security System and is financed through employer- and employee-paid taxes. The level of DI benefits received is determined using a formula based on an individual's

earnings history. The proportion of income that is replaced decreases as past earnings increase so that low-wage workers replace a larger fraction of their earnings than do high-wage workers.

The disability determination process involves multiple steps, as diagrammed in Figure 1. The first step is the submission of an initial application to the Social Security Administration office for the Disability Determination Stage (DDS) review. If the applicant meets the non-medical criteria (such as age and prior employment requirements), disability examiners and medical staff assess written medical evidence regarding the applicant's ability to perform work-related activities. Examiners take into account health status, age, education, and work experience as well as the transferability of the applicant's skills. If the disability examiner concludes that the applicant cannot be expected to engage in any substantial gainful activity, a disability award is made. Partial disability awards can also be made. Approximately 75% of claims are awarded at the DDS review. Cases that are more difficult to judge (such as mental illness and low back pain) are often denied at this step.

If the DI claim is denied at the DDS review, the individual may appeal the decision within 2 months to the Court of Appeals. About 25% of all denials are appealed. DI appeals are reviewed by Administrative Law Judges (ALJs). The ALJ must consider the application using the same criteria as the initial determination, but the applicant may present new information in writing. Judges can either allow a case, deny a case, or issue a remand (which means the case is sent back to the DDS Review stage to be re-evaluated with updated information).[6] Approximately 15% of all claims that were appealed are allowed at the ALJ level. If the case is denied at the ALJ level, the applicant can always choose to start a new DI case by re-applying to the DDS Review stage. Seventy-five percent of denied appellants eventually reapply, with 65 percent of these being ultimately allowed DI. If a case is denied at the ALJ level, it can also be appealed to the higher courts, but very few individuals exercise this option.

### Random assignment of DI cases to judges

In Norway, the hearing of appeals is centralized in Oslo, where cases are handled for the entire country. Prior to 1998, there was only one department. Afterwards, there were four equally-sized departments; however, there is no specialization in the four departments and all judges are housed in the same building. Within each department, the assignment of a case to an Administrative Law Judge is done by the department head without knowing the content of the case, as stipulated in the rules set forth for the Administrative Law Court since its inception in 1967. The rules state that assignment should be done "by the drawing of lots." In practice, cases are assigned on a rotating basis depending on the date they are received and the alphabetical ordering of a judge's last name.[7]

[6] Average processing time at the DDS stage is 6 months, while average processing time at the appeal stage is 4 months. Remands are uncommon, accounting for only 5 percent of appeal outcomes. In our baseline analysis, we code remanded cases as rejected. As a robustness check, we coded remanded cases as allowed or denied based on their eventual outcome after they are reconsidered by the DDS case worker with updated information and the results are similar.

[7] We verified these rules with the current Head of the Administrative Law Court, Knut Brofoss. The rules are explained in "Veileder for Saksbehandlingen i Trygderetten" (Guidelines for Processing Cases in the Court of Appeals). We have also presented our paper at internal seminars with the current set of judges and department heads to make sure that we have understood how the cases are handled and assessed.

Our setting has several attractive features: (i) the handling of cases is centralized in one location, (ii) judges do not specialize by medical condition, region of country, or other aspects of the case, (iii) the judge assesses the written evidence on the appellant's case; there is never any personal contact between the judge and those who appeal, and (iv) an individual cannot choose an alternate judge after being assigned a judge.

A key to our design is not only that the assignment of judges is random, but also that some judges are more lenient than others. We measure judge leniency based on the average allowance rate in all other cases a judge has handled. This measure is based on all the cases a judge has ever handled, and not just those cases appearing in our estimation sample. On average, the 76 judges in our sample have handled a total of 375 cases. To construct the judge leniency measure, we calculate the leave-out mean judge allowance rate and regress this measure on fully interacted year and department dummies; this is because the randomization occurs among the pool of judges within each department. We use the residual from this regression as our judge leniency measure. This approach controls for any differences over time or across departments in the quality of applicants and the leniency of the judges.

**Verifying random assignment**

Table 1 empirically verifies that the hearing office complied with the random allocation procedure. This table conducts the same type of statistical test that would be done for an actual experiment to verify compliance with randomization. We find strong empirical support for the claim that the DI system in Norway randomly assigns judges to individuals who appeal their cases, conditional on fully interacted year and department dummies. The first column documents that demographic, work and health variables are highly predictive of whether an appealed case will be allowed. Column 3 examines whether our measure of judge leniency can be predicted by these same characteristics. Even though the set of characteristics are highly predictive of case outcomes, they are not statistically related to the leniency of the judge assigned to a case: none of the 16 variables are statistically significant at the 10% significance level and the variables are not jointly significant either. In fact, the point estimates are close to zero, and taken together, the variables explain only 0.1 percent of the variation in our measure of judge leniency. Note in particular the insignificance of the disorder variables. This is consistent with the lack of specialization by type of disability in Norway, something which is not true in many other countries.

A natural question is why some judges are more lenient than others. While we do not have detailed characteristics of the judges, we do know the number of cases they have handled. Whereas experienced judges appear to be slightly less lenient, experience accounts for only a small fraction of the total variation in allowance rates across judges (see Appendix Figure A.1). Other unobserved factors must be driving the underlying variation. It is important to recognize that as long as judges are randomly assigned, it does not matter why some judges are more lenient than others.

### Instrument and empirical model

We use variation in DI allowance generated from the random assignment of appeal judges as an instrument to estimate the economic consequences of DI allowance. We estimate judge leniency by taking the average allowance rate in all other cases a judge has ever handled, adjusted for year and department effects, as we did for Table 1.[8] As we document below, some judges are systematically more lenient than others, which gives exogenous variation in the probability an individual is allowed DI in the appeals process.

Our baseline empirical model can be described by the following two-equation system:

$$A = \alpha + \gamma Z + \theta X + v \tag{1}$$

$$Y = \mu + \beta A + \lambda X + u \tag{2}$$

where $Z$ denotes a judge's leniency, $A$ is an indicator for whether an individual is allowed DI in the appeal process, $X$ is a vector of control variables, and $Y$ is a dependent variables of interest that is measured at some point after the allowance decision (e.g. earnings three years after the decision).

We perform 2SLS with equation (1) as the first stage and equation (2) as the second stage, with the goal of consistently estimating the parameter $\beta$. Note that our dependent variable in the first stage equation is whether the individual is allowed DI in the appeal process, rather than whether he participates on DI at the time when $Y$ is measured. Our specification of the first stage alleviates concerns about the exclusion restriction: the 2SLS estimate of $\beta$ captures any effect which operates through whether the individual is allowed DI in the appeal process, including participation in DI, subsequent reapplications to the DI program after being denied or any other causal change in parental behavior. We can also estimate the reduced form effect by directly regressing $Y$ on $Z$ and $X$.

# 3 Data and Background

## 3.1 Data and Sample Restrictions

Our analysis employs several data sources that we can link through unique identifiers for each individual. Information on DI benefits comes from social security registers that contain complete records for all individuals who entered the DI program during the period 1967-2010. The data set includes information on the individual's work history and medical diagnosis, the month when DI was awarded (or denied), and the level of DI benefits received. We link this information with administrative data from the hearing office on all appeals from 1989 to 2011. The data set contains information on dates of appeal and decision, the outcome of the appeal, and unique identifiers for both judges and applicants. We merge these data sets with administrative registers provided by Statistics Norway, using a rich longitudinal database that covers every resident from 1967 to 2010. For each year, it contains individual demographic information (including sex, age, and education).

[8] Although the instrument is pre-estimated, there is no need to adjust the standard errors of the IV estimates; such adjustments are necessary with generated regressors but not with generated instruments.

Since 1993, we also have detailed information on all sources of annual income (such as earnings, income from self employment, capital income, cash transfers) from tax records of each individual. The data contains unique identifiers that allow us to match spouses and parents to children.

The coverage and reliability of Norwegian registry data are rated as exceptional in international quality assessments (see e.g. Atkinson *et al.* , 1995). Importantly, the Norwegian income data has several advantages over those available in most other countries. First, there is no attrition from the original sample because of the need to ask permission from individuals to access their tax records. In Norway, these records are in the public domain. Second, our income data pertains to all individuals, and not only to jobs covered by social security. Third, we have data on all sources of income, without any top or bottom coding.

Our empirical analysis considers individuals who appeal an initially denied DI claim.[9] To observe individuals for at least 4 years after the appeal decision, our estimation sample consists of individuals whose appeal decision was made during the period 1993-2005. Following Maestas *et al.* (2013) and French & Song (2013), our baseline estimation excludes observations for which the assigned appeal judge has handled few cases (less than ten during the period 1989 to 2011). The reason for this sample restriction is to reduce the noise in our instrument. We also exclude individuals who were older than 62 years at the time he or she appealed. The reason for this age restriction is to avoid program substitution between DI and early retirement schemes.

In Table 2, we document the key characteristics of the sample of individuals who apply for DI and our baseline sample of individuals who appeal an initially denied DI claim. Those who appeal are on average more likely to be female, less educated and foreign born, and have lower prior earnings compared to the group of initial applicants. Sixty-two percent of applicants claim mental or musculoskeletal disorders, whereas this figure is 71 percent for appellants.

### 3.2 Institutional Background

There are a number of similarities and a few key differences between the DI systems in the U.S. and in Norway (see Autor & Duggan, 2006; Kostol & Mogstad, 2014). In both countries, DI is one of the largest transfer programs. However, the incidence of receipt of DI benefits is lower in the U.S. than in Norway. Figure 2 shows this distinction by displaying the evolution of DI in the two countries. Whereas the rate of DI receipt in a given year is consistently higher in Norway than in the U.S., the time trends are quite similar.[10] From 1961 to 2012, the rate of receipt increased from 2.2 to 9.7 percent in Norway and from 0.8 to 5.4 percent in the U.S. While Norway's rate has leveled off at about 10 percent in recent years, the U.S. DI rate continues to rise and is projected to exceed

[9]Some individuals have several denied DI claims over the period we consider. In such cases, we restrict our sample to the individual's first denied DI claim.

[10]The cross-country difference in DI coverage is unlikely to explain the entire discrepancy in the incidence of DI: although virtually all non-elderly adults are covered in Norway, more than 80 percent of all non-elderly adults are covered in the U.S. The remaining difference could be a function of underlying differences in screening stringency, the generosity of the programs or the frequency with which people apply for disability benefits. Milligan & Wise (2011) argue that differences in health are unlikely to explain much of the observed differences in DI rates across developed countries.

7 percent by 2018 (Burkhauser & Daly, 2012).

In both countries, the expansion of the DI rolls in recent decades appears to be driven by the liberalization of the screening process, which led to a rapid increase in the share of DI recipients suffering from difficult-to-verify disorders such as mental illness and musculoskeletal disease.[11] Because these are early-onset disorders with low mortality at young ages, DI recipients with such diagnoses tend to participate in the program for relatively long periods. As a result, the DI exit rates in both countries have decreased in the last few decades, with progressively fewer DI recipients reaching retirement age or dying in a given year (see Appendix Figures A.2 and A.3).

There are a few noticeable differences between the two countries. DI recipients in Norway tend to be older and have slightly higher earnings prior to a disability award. One possible explanation for this is that the U.S. SSDI program is less generous.[12] The differences in characteristics are, however, less pronounced than one might expect. For instance, almost 60 percent of DI recipients suffer from difficult-to-verify disorders (mental illness and musculoskeletal disorders) in both the U.S. and Norway (see Appendix Table A.1).

Another difference is that the appeal process plays a more important role in the U.S. than in Norway. While 48 percent of the initially rejected applicants appeal in the U.S. (French & Song, 2013), only 25 percent of the initially rejected appeal in Norway. Success rates for appeals are much higher in the U.S. Appendix Table A.2 compares the characteristics of individuals who apply for DI and those who appeal an initially denied DI claim in the two countries. In both the U.S. and Norway, appellants are more likely to be younger, less connected to the labor market, and more likely to suffer from difficult-to-verify disorders, as compared to the the initial group of applicants.

# 4 Assessment of the instrument

We begin our presentation of results by providing evidence on the relevance and validity of the instrument.

## 4.1 Instrument relevance

Figure 3 provides a graphical representation of the first stage of our IV model. In the background of this figure is a histogram for the density of judge leniency, which captures the average judge allowance rate in the other cases a judge has handled. We note the judge leniency measure is

[11] See Autor & Duggan (2006) for a discussion of this phenomenon. In the U.S., the 1984 congressional reforms shifted the focus of screening from medical to functional criteria. In Norway, the medical eligibility criteria were relaxed earlier and more gradually.

[12] For a typical DI recipient in Norway, Kostol & Mogstad (2014) calculate the replacement rate would be 31 percent according to U.S. program rules and 58 percent according to Norwegian program rules. Factoring in health insurance coverage increases the effective replacement rate to over 50 percent in the U.S. In Norway, all citizens are eligible for health insurance through the Social Insurance system. These calculations abstract from differences before an individual goes on to DI. In Norway, most individuals must first participate in the sick leave program which lasts at most one year and precludes full-time work. This is an departure from the U.S., where eligibility requires an impairment that is expected to last at least a year but does not require participation in any programs before DI receipt.

calculated from all cases the judge has ever handled, not just the cases in our estimation sample. On average, each judge has handled a total of 375 cases. The mean of the leniency variable is .15 with a standard deviation of .05. The histogram reveals a wide spread in judge leniency, with approximately 18% of cases allowed by a judge at the 90th percentile compared to approximately 8% at the 10th percentile.

Figure 3 also graphs the relationship between judge leniency and the appellant's allowance rate. The graph is a flexible analog to the first stage equation (1), where we plot a local linear regression of actual individual allowance against judge leniency. The individual allowance rate is monotonically increasing in our leniency measure, and is close to linear. A 10 percentage point increase in the judge's allowance rate in other cases is associated with an approximately 8 percentage point increase in the probability the individual's case is allowed.

In Table 3, we turn to a regression based analysis. We include fully interacted year and department dummies in Panel A, but otherwise include no other controls. In each column, we regress a dummy variable for whether an individual is allowed DI at the appeal stage on our judge leniency measure. The only reason the first stage estimates vary with years since court decision is sample attrition due to death or emigration. Because there is very little attrition, the first stage estimates barely move across the columns. As a result, any variation in the IV estimates by years since court decision can be attributed to changes in the reduced form impact of judge leniency on the outcomes of interest.

### 4.2 Instrument validity

In order for judge leniency to be a valid instrument, appellants' assignment to judges must be uncorrelated with case characteristics. Table 1 provided strong empirical support for the claim that the DI system in Norway randomly assigns appeal judges within each department and year. As a second test, Panel B of Table 3 explores what happens if a large set of control variables for observable characteristics are added to the baseline regressions. If judges are randomly assigned, the addition of these control variables should not significantly change the estimates, as the case characteristics should be uncorrelated with judge leniency. As expected, the coefficients do not change appreciably.

While random assignment of cases to judges is sufficient for a causal interpretation of the reduced form estimates, the IV estimates require two additional assumptions. The first is that the leniency of the individual's judge affects the outcome of interest only through the individual's allowance decision, and not directly in any other way. One attractive feature of the process in Norway makes this exclusion restriction likely to hold: the appeal is presented in writing, so there is never any personal contact between the judge and those who appeal. What individuals (and their families) observe is the allowance or denial decision of the judge.

A possible caveat is that appeal processing time could differ systematically by the leniency of the judge (see e.g. Autor (2011b)) and that this could directly affect the outcomes of interest. To examine this, we calculated a judge's average processing time based on the residual average processing time in the other cases a judge has handled after controlling for a fully interacted set

of time and department dummies in a regression. Panel C of Table 3 shows that the first stage estimates do not change appreciably if we control for characteristics other than judge leniency that we observe in our data, such as his average processing time.

The final assumption needed for a causal interpretation of the IV estimates is monotonicity of judges' appeal decisions. In our setting, the monotonicity assumption is that cases allowed by a strict judge would also have been allowed by a more lenient judge, and similarly that cases denied by a lenient judge would also have been denied by a stricter judge. One testable implication of the monotonicity assumption is that the first stage estimates should be non-negative for all subsamples. Indeed, when separately estimating the first stages based on the observable characteristics of the individual, the estimates are consistently positive and sizable, in line with the monotonicity assumption.[13]

# 5 Main results

## 5.1 Labor outcomes and DI benefits

Table 4 reports reduced form and IV estimates for labor outcomes and DI benefits. These results are useful in understanding not only the behavioral responses of individuals to the appeal decision, but also as background for understanding the effect of DI allowance on the participation rates in other social programs and its impact on family disposable income.

As in Table 3, we separately estimate the effects for each of the four years after the appeal decision in which we have a balanced sample of individuals. In every panel, we report reduced form estimates which regress the outcome of interest on our judge leniency measure; we also scale these reduced forms by the first stages, which gives us the IV estimates of the effect of DI allowance at appeal. Every specification includes fully interacted year and department dummies as well as controls for the observable case characteristics.

In Panel A, we consider the impact on DI participation over a four-year period after the appeal decision. Our results show that many individuals who are initially denied re-apply and are eventually allowed DI. When looking four years after the appeal decision, the first row of this panel shows a reduced form estimate of .4. This estimate implies that when judge leniency rises by 10 percentage points, an individual's DI participation will rise by roughly 4 percentage points after four years. The next row takes the reduced form estimate and divides it by the first stage estimate. For example, four years after the court decision, being allowed DI at the appeal increases the probability of DI participation by 52 percentage points.

Panel B of Table 4 displays results for labor market participation (defined as having positive earnings). The results show fairly small but significant effects on the probability of working. For example, four years after the court decision, being denied DI at the appeal increases the probability of labor force participation by 16 percentage points.

[13] The results are available from the authors upon request. See also Dahl *et al.* (2013).

Panels C and D complement by displaying results for earnings and payments of DI benefits. We find that being allowed DI at the appeal leads to a large increase in benefit payments over the four year period. However, the effects decline over time as an increasing number of initially denied individuals re-apply and are eventually allowed DI. By way of comparison, the effects on earnings are smaller in magnitude and decline only modestly over time.

Our IV estimates can be interpreted as local average treatment effects (LATE) for the individuals who comply with the instrument, and therefore could have received a different allowance decision had their case been assigned to a different judge. A natural question is: Over this four-year period, how many of these compliers would be working and how much DI benefits would they collect if they had been denied at the appeal. As shown in Dahl *et al.* (2013), we can recover these potential outcome levels by combining (i) the shares of never takers and compliers to the instrument with (ii) estimates of the mean outcomes of individuals who were not allowed with the most lenient or strictest judges. In Figure 4, we do these calculations and decompose the LATE into the potential outcome of compliers if they had been denied at the appeal and their potential outcome if they had been allowed. Because of the random assignment of judges, the potential outcomes are the same for the two groups in the years prior to the allowance decision.

The graphs in Figure 4 tell an interesting story: The labor outcomes of denied compliers change little over time, even though many re-apply and are eventually allowed DI. This suggests there are two distinct types of denied appellants among the compliers: A small but non-negligible group who return to work and earn, on average, about $5,000 per year; a larger group who instead of working continue to apply for DI and potentially other transfer programs.

### 5.2 Benefit substitution

Table 5 reports reduced form and IV estimates for total transfers (DI benefits + all other cash transfers), other cash transfers than DI benefits, and payments from the social assistance program (traditional welfare). The social assistance program is a stigmatized, last-resort safety net and there are no clear rules regarding eligibility or benefit amounts, with discretion being left to the local social worker. For comparison, we also repeat the estimated effects on DI benefit payments.

The results in Table 5 point the importance of taking benefit substitution into account when making statements about the consequences of tightening disability screening, both from the perspective of public finances and in terms of household income or welfare. In particular, the impact of being allowed at the appeal is considerably smaller for total transfers than for DI benefits. This is because individuals who are denied apply for and get on other transfer programs. However, traditional welfare like social assistance does not appear to be a close substitute to DI participation; instead, participation in vocational training or medical rehabilitation programs seems to be driving the benefit substitution.

In Figure 5, we decompose these LATEs into potential outcome levels of compliers if they had been denied at the appeal and if they had been allowed. The graphs reveal that total transfer payments change little if individuals are denied at appeal; what changes is the transfer payments to

individuals who are allowed. We can also see that individuals who are denied start collecting other benefits for a few years after the appeal decision, before getting onto the DI program.

### 5.3 Household disposable income

Table 6 presents the final set of regression results, where we look at earnings, total transfers and disposable income at the household level (per capita). By doing so, we incorporate that spouses may also respond to the allowance decision, either by changing their labor supply or by applying for their own benefits. Figure 6 complements by showing the potential levels of these household outcomes for compliers if they had been denied at the appeal and if they had been allowed.

By comparing the results for individual earnings in Table 4 to the household results in Panel A of Table 6, it is clear spouses of denied appellants respond by working more. As a result, household earnings increases significantly if an appellant is denied DI. However, some of the gains in earnings of the spouses are offset by a decline in the transfer payments they receive; this is evident from the relatively large effect on total transfers to the household in panel B of Table 6, as compared to the effect on total transfers to the appellants in Panel A of Table 5.

Taken together, the increase in transfer payments exceed the decline in earnings, so that household disposable income rises significantly if the individual is allowed DI at the appeal. As shown in Panel C of Table 6, the drop in disposable income is relatively large. When looking four years after the appeal, for example, the drop amounts to approximately 20 percent of the sample average in household income. This findings shows that, even in a setting with substantial benefit substitution and significant labor supply response among spouses, making the disability screening stricter would substantially reduce the income available for household consumption.

## 6 Conclusion

Two key questions in thinking about policy proposals to tighten the screening process are to what extent the DI program discourages work, and how valuable the insurance is to individuals and families. On the one hand, stricter screening criteria could reduce the number of individuals who still have substantial work capacity from collecting benefits. On the other hand, it could result in income or welfare loss to individuals and families who value the insurance aspect of the program in cases of true disability.

Assessing this insurance-incentive tradeoff in tightening the disability screening has proven difficult for several reasons. It is often difficult to access data on disposable income or consumption which can be linked with social security records. Another key challenge is that individuals receiving DI are likely to differ from non-recipients in important ways, both in observable and unobservable dimensions. To date, most empirical research focus on the labor market impact of DI receipt, and less is known about how a tightening of the disability screening would affect individuals and families on other margins.

This paper helps to close that gap by analyzing the broader economic consequences of tightening

the disability screening. The key to our research design is twofold. First, we are able to link a number of administrative registers from Norway, providing a unique population panel data set with detailed information about every individual and family. Second, the DI system in Norway randomly assigns judges to DI applicants whose cases are initially denied; some appeal judges are systematically more lenient, which leads to random variation in the probability an individual will get onto DI. We used the random assignment of appeal judges to estimate the causal effects of DI allowance on a variety of outcomes, including not only individual labor force participation and earnings but also participation in other social insurance programs and family disposable income. In ongoing work, we are collecting additional data on assets and consumption, which would allow us to estimate the degree of insurance provided by the DI program, over and above self-insurance through savings and the spouse's labor supply.

## References


Aizer, Anna, & Joseph J. Doyle, Jr. 2013 (June). *Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly-Assigned Judges.* Working Paper 19102. National Bureau of Economic Research.

Atkinson, A.B., Rainwater, L., & Smeeding, T.M. 1995. Income Distributions in OECD countries: evidence from the Luxembourg Income Study. *OECD Publications and Information Center.*

Autor, D. H., & Duggan, M. G. 2003. The rise in the disability rolls and the decline in unemployment. *Quarterly Journal of Economics*, **118**(1), 157–205.

Autor, D. H., & Duggan, M. G. 2006. The growth in the social security disability rolls: A fiscal crisis unfolding. *Journal of Economic Perspectives*, **20**(3), 71–96.

Autor, D. H., Levy, F., & Murnane, R. J. 2003. The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, **118**(4), 1279–1333.

Autor, David. 2011a. The Unsustainable Rise of the Disability Rolls in the United States: Causes, Consequences, and Policy Options. *NBER*, **No. 17697**.

Autor, David. 2011b. The Unsustainable Rise of the Disability Rolls in the United States: Causes, Consequences, and Policy Options. *NBER Working Paper*, **No. 17697**.

Borghans, Lex, Gielen, Anne, & Luttmer, Erzo. 2012. Social Support Substitution and the Earnings Rebound: Evidence from a Regression Discontinuity in Disability Insurance Reform. *NBER Working Paper*, **No. 18261**.

Bound, J. 1989. The Health and Earnings of Rejected Disability Insurance Applicants. *American Economic Review*, **79**(3), 482–503.

Burkhauser, Richard V., & Daly, Mary C. 2012. Social Security Disability Insurance: Time for Fundamental Change. *Journal of Policy Analysis and Management*, **31**(2), 454–461.

Campolieti, Michele, & Riddell, Chris. 2012. Disability policy and the labor market: Evidence from a natural experiment in Canada, 1998-2006. *Journal of Public Economics*, **96**, 306–316.

Dahl, Gordon B., Kostol, Andreas Ravndal, & Mogstad, Magne. 2013 (July). *Family Welfare Cultures.* Working Paper 19237. National Bureau of Economic Research.

De Jong, Philip, Lindeboom, Maarten, & Van der Klaauw, Bas. 2011. Screening disability insurance applications. *Journal of the European Economic Association*, **9**(1), 106–129.

Dobbie, Will, & Song, Jae. 2013. Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection. *Manuscript.*

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

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

French, Eric, & Song, Jae. 2013. The Effect of Disability Insurance Receipt on Labor Supply, Appeals and Re-Applications: Evidence from Hearing Office Level Decisions. *American Economic Journal: Applied*, **forthcoming**.

Gruber, J. 2000. Disability insurance benefits and labor supply. *Journal of Political Economy*, **108**(6), 1162–1183.

Karlstrom, Anders, Palme, Marten, & Svensson, Ingemar. 2008. The employment effect of stricter rules for eligibility for DI: Evidence from a natural experiment in Sweden. *Journal of Public Economics*, **92**, 2071–2082.

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

Kostol, Andreas Ravndal, & Mogstad, Magne. 2014. How Financial Incentives Induce Disability Insurance Recipients to Return to Work. *American Economic Review*, **104**(2), 624–55.

Low, Hamish, & Pistaferri, Luigi. 2012. Disability Insurance and the Dynamics of the Incentive-Insurance tradeoff. *NBER Working Paper*.

Maestas, Nicole, Mullen, Kathleen J., & Strand, Alexander. 2013. Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. *American Economic Review*, **103**(5), 1797–1829.

Meyer, Bruce D., & Mok, Wallace K.C. 2013 (March). *Disability, Earnings, Income and Consumption*. Working Paper 18869. National Bureau of Economic Research.

Milligan, K. S., & Wise, D. A. 2011. Social Security and Retirement around the World: Mortality, Health, Employment and Disability Insurance Participation and Reforms. *Harvard Kennedy School Working paper*.

Moore, T. J. 2011. The Employment Effects of Terminating Disability Benefits: Insights from Removing Drug Addictions as Disabling Conditions. *Working Paper University of Maryland*.

OASDI-Trustees-Report. 2012. *The 2012 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds*. Tech. rept.

OECD. 2010. *Sickness, Disability and Work: Breaking the Barriers. A synthesis of findings across OECD countries*. OECD Publishing.

Parsons, Donald O. 1991. Self-Screening in Targeted Public Transfer Programs. *Journal of Political Economy*, **99**(4), 859–876.

Staubli, Stefan. 2011. The impact of stricter criteria for disability insurance on labor force participation. *Journal of Public Economics*, **95**(9), 1223–1235.

von Wachter, Till, Song, Jae, & Manchester, Joyce. 2011. Trends in Employment and Earnings of Allowed and Rejected Applicants to the Social Security Disability Insurance Program. *The American Economic Review*, **101**(7), 3308–29.

# Tables and Figures

Figure 1: **DI Application and Appeals Process.**

*Notes:* This figure summarizes the description of the DI application and appeal process in subsection 2. Table 2 provides the summary statistics for all applicants in the period 1992-2003.

Figure 2: **Trends in DI Receipt in Norway and the U.S.**


*Notes:* This figure displays trends in DI receipt in Norway and the U.S. (see Section 2).. U.S. trends are based on (Autor & Duggan, 2006) for 1957-2005 and SSA Office of the Chief Actuary for 2006-2012. Norwegian trends are based on SSA Statistical Supplements. Incidence of DI receipt defined as the percent of .the relevant adult population receiving DI benefits (age 18-67 in Norway; age 25-64 in the US).

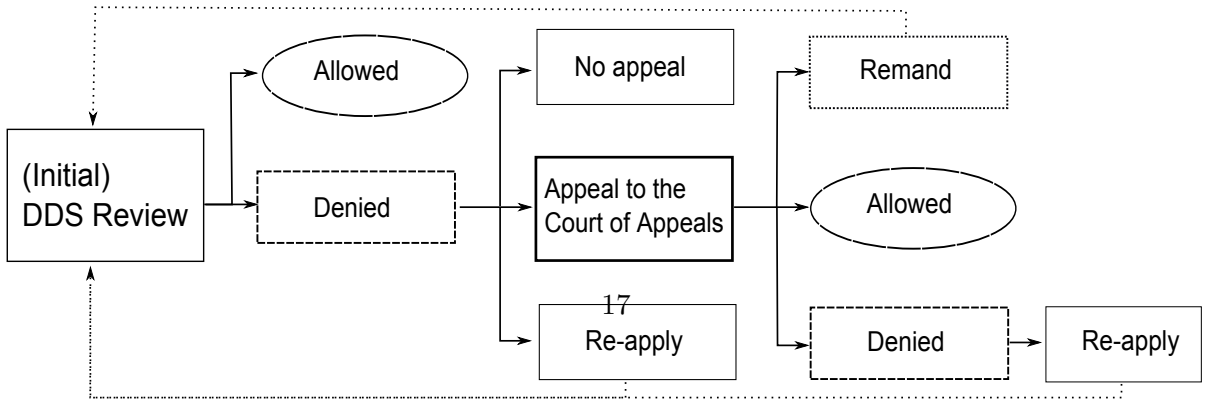
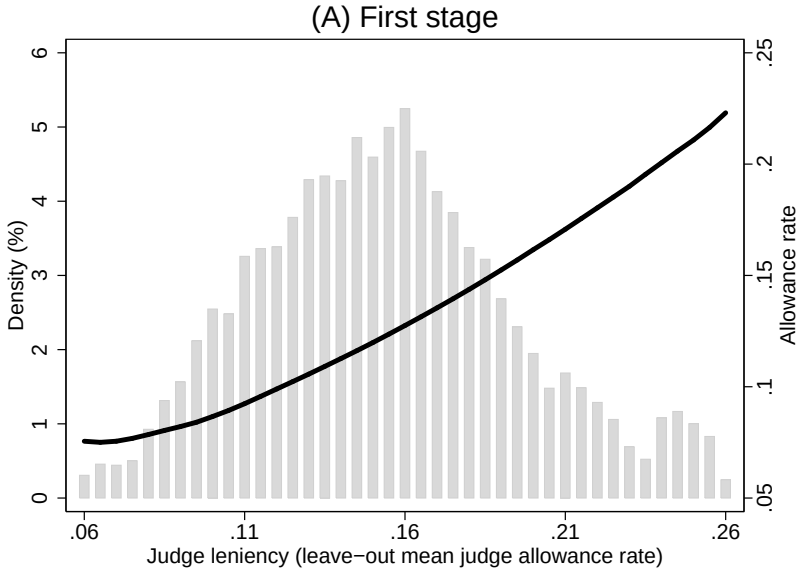
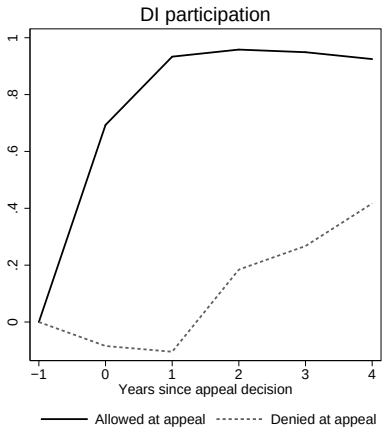


Figure 3: **Effect of Judge Leniency on DI allowance.**

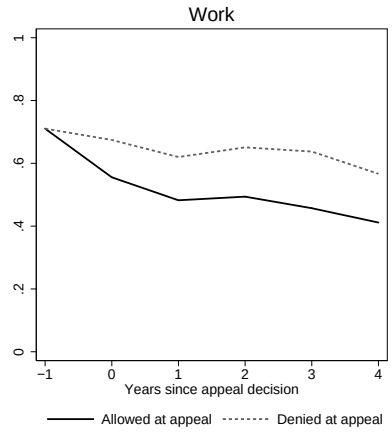



*Notes:* This figure displays the effect of judge leniency on DI allowance. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1993-2005 (see Section 3 for further details). There are 76 different judges. The solid lines are a local linear regressions of allowance on judge leniency. The histogram of judge leniency is shown in the background of both figures (top and bottom 1% excluded from the graph).

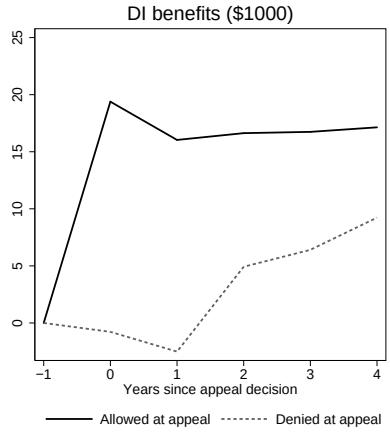
Figure 4: **Potential Outcomes: Labor outcomes and DI benefits.**



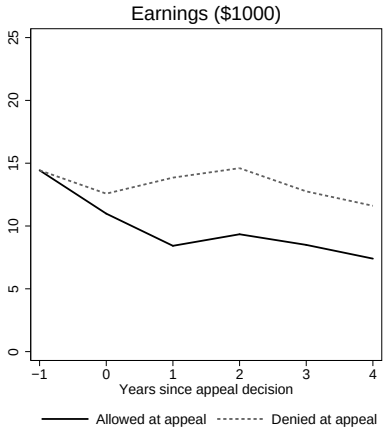

(a) DI participation




(b) Labor market participation



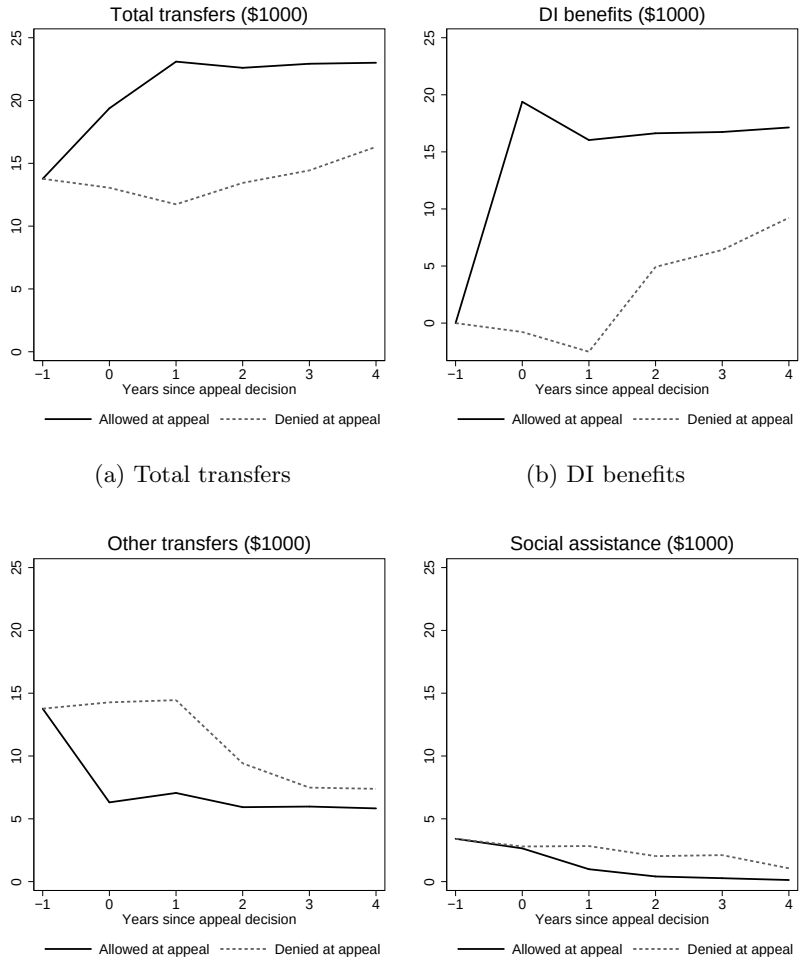

(c) DI benefits




(d) Labor market earnings

*Notes:* These figures display the decomposition of our LATE estimates into potential outcomes for allowed and denied complier appellants (see Dahl *et al.* (2013) for details).

Figure 5: **Potential Outcomes: Benefit substitution,**




(a) Total transfers

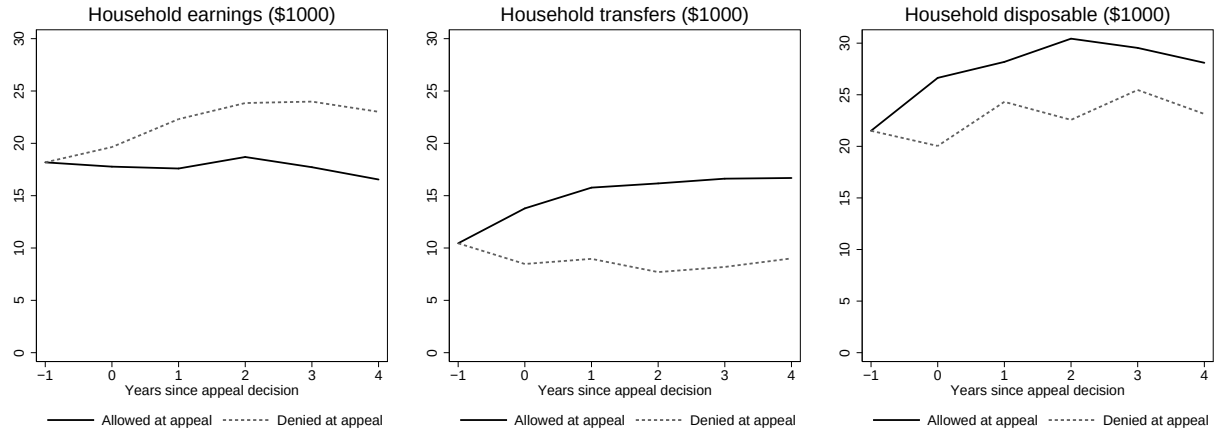
(b) DI benefits

(c) Other transfers

(d) Social assistance

*Notes:* These figures display the decomposition of our LATE estimates into potential outcomes for allowed and denied complier appellants (see Dahl *et al.* (2013) for details).

Figure 6: **Potential Outcomes: Household outcomes (per capita).**




(a) Earnings (b) Total transfers (c) Disposable income

*Notes:* These figures display the decomposition of our LATE estimates into potential outcomes for allowed and denied complier appellants (see Dahl *et al.* (2013) for details).

Table 1: **Testing for Random Assignment of Cases to Judges.**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable: | **Case Allowed** | | **Judge Leniency** | |
| | coeff. | s.e. | coeff. | s.e. |
| Age (at the time of decision) | 0.0043*** | (0.0003) | 0.0001 | (0.0001) |
| Number of persons in household | -0.0152*** | (0.0018) | -0.0002 | (0.0003) |
| Average indexed earnings | 0.0010*** | (0.0002) | 0.0000 | (0.0000) |
| Female | 0.0205*** | (0.0055) | 0.0010 | (0.0012) |
| Married | 0.0161*** | (0.0066) | 0.0012 | (0.0012) |
| Foreign born | -0.0482*** | (0.0082) | 0.0004 | (0.0015) |
| Less than high school degree | -0.0229*** | (0.0060) | -0.0006 | (0.0008) |
| High school degree | 0.0196*** | (0.0056) | 0.0005 | (0.0008) |
| Any college | 0.0112 | (0.0111) | 0.0005 | (0.0014) |
| Children below age 18 | -0.0597*** | (0.0048) | -0.0011 | (0.0009) |
| Musculoskeletal disorders | -0.0162*** | (0.0059) | 0.0003 | (0.0017) |
| Mental disorders | 0.0080 | (0.0075) | -0.0008 | (0.0025) |
| Circulatory system | 0.0259 | (0.0159) | -0.0002 | (0.0022) |
| Respiratory system | -0.0147 | (0.0147) | -0.0028 | (0.0020) |
| Neurological system | 0.0498*** | (0.0210) | 0.0022 | (0.0021) |
| Endocrine diseases | 0.0436*** | (0.0180) | -0.0027 | (0.0032) |
| F-statistic for joint significance | 26.67 | | 1.3 | |
| [p-value] | [.001] | | [.22] | |
| N | 13,951 | | 13,951 | |

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

*Notes:* This table displays the test of whether the hearing office complied with the random allocation procedure described in Section 2. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1993-2005 (see Section 3 for further details). There are 76 different judges. Columns 1 and 3 display OLS estimates from separate regressions of whether a case is allowed or judge leniency, respectively, on appellant characteristics. F-statistics are obtained from OLS estimation on the combined set of applicant characteristics. All regressions are done by double residual regressions, by first purging the leniency measure for year of appeal fixed effects. Characteristics of appellants are measured prior to the appeal. Average indexed earnings is mean earnings for the last ten years prior to appeal, children is equal to 1 if appellant has children under age 18 and 0 otherwise, and any college is equal to one if a person has some college or has a college degree. Body system codes are based on ICD-10 diagnostic codes.

Table 2: **Descriptive Statistics.**

| | DI applicants | | DI appellants | |
|---|---|---|---|---|
| **Characteristic** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** |
| Age (at the time of decision) | 48.52 | [10.04] | 46.51 | [9.32] |
| Number of persons in household | 4.43 | [2.39] | 2.87 | [1.59] |
| Average indexed earnings | 33.07 | [23.55] | 21.50 | [19.84] |
| Female | 0.56 | [0.50] | 0.64 | [0.48] |
| Married | 0.57 | [0.49] | 0.57 | [0.49] |
| Foreign born | 0.08 | [0.27] | 0.18 | [0.38] |
| Less than high school degree | 0.44 | [0.50] | 0.50 | [0.50] |
| High school degree | 0.42 | [0.49] | 0.39 | [0.49] |
| Any college | 0.13 | [0.33] | 0.11 | [0.31] |
| Children below age 18 | 0.30 | [0.46] | 0.44 | [0.50] |
| Musculoskeletal disorders | 0.37 | [0.48] | 0.45 | [0.50] |
| Mental disorders | 0.25 | [0.44] | 0.26 | [0.44] |
| Circulatory system | 0.08 | [0.28] | 0.04 | [0.19] |
| Respiratory system | 0.03 | [0.17] | 0.03 | [0.16] |
| Neurological system | 0.05 | [0.23] | 0.04 | [0.19] |
| Endocrine diseases | 0.02 | [0.14] | 0.04 | [0.20] |
| DI allowed | 0.78 | [0.41] | 0.13 | [0.33] |
| Number of observations | 254,196 | | 13,951 | |

Standard deviations [in square brackets]

*Notes:* This table displays means for applicants and appellants described in Section 2. The applicant sample consists of all claims made during the period 1992-2003 by individuals who are at most 62 years of age. The sample of appellants (see section 3) is a subgroup of the denied applicant sample. Unless otherwise stated, all characteristics are measured the year before application/appeal. Average indexed earnings is mean earnings for the last ten years prior to appeal, children is equal to 1 if appellant has children under age 18 and 0 otherwise, and any college is equal to one if a person has some college or has a college degree. Body system codes are based on ICD-10 diagnostic codes. Nominal values are deflated to 2005 and represented in US dollars using the average exchange rate NOK/$ = 6.

Table 3: **First Stage**: **Judge leniency and DI allowance**.

| **Years after decision** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| **Panel A: No covariates** | | | | |
| Judge leniency | 0.788*** | 0.791*** | 0.793*** | 0.800*** |
| | (0.084) | (0.085) | (0.084) | (0.083) |
| **Panel B: With individual covariates** | | | | |
| Judge leniency | 0.760*** | 0.760*** | 0.764*** | 0.769*** |
| | (0.081) | (0.082) | (0.081) | (0.081) |
| **Panel C: With judge characteristics** | | | | |
| Judge leniency | 0.781*** | 0.784*** | 0.792*** | 0.801*** |
| | (0.076) | (0.077) | (0.077) | (0.078) |
| Dependent mean | 0.126 | 0.126 | 0.126 | 0.126 |
| Obs | 14,311 | 14,180 | 14,039 | 13,932 |

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Number of Judges: 76.

*Note:* This table displays the first stage coefficients of equation 1. In Panel A, DI allowance is regressed on judge leniency and fully interacted year of appeal and department dummies. In Panel B, we include flexible controls for individual characteristics as described in Table 1, as well as month of appeal and county fixed effects. In Panel C, we include judge characteristics (judge experience, judge leave-out-mean remand rate and processing time).

Table 4: **Main Results I: Labor outcomes and DI benefits.**

| **Years after decision** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| **Panel A.** | | **DI participation** | | |
| Judge leniency | 0.790*** | 0.592*** | 0.529*** | 0.401*** |
| | (0.068) | (0.080) | (0.072) | (0.065) |
| Allowed DI | 1.036*** | 0.777*** | 0.690*** | 0.520*** |
| | (0.085) | (0.117) | (0.107) | (0.097) |
| Dependent mean | 0.302 | 0.430 | 0.517 | 0.575 |
| **Panel B.** | | **Labor market participation** | | |
| Judge leniency | -0.106* | -0.119* | -0.134** | -0.124** |
| | (0.060) | (0.065) | (0.061) | (0.055) |
| Allowed DI | -0.140* | -0.156* | -0.176** | -0.161** |
| | (0.079) | (0.085) | (0.082) | (0.070) |
| Dependent mean | 0.591 | 0.578 | 0.561 | 0.539 |
| **Panel C.** | | **DI benefits ($1000)** | | |
| Judge leniency | 14.123*** | 9.151*** | 8.120*** | 6.358*** |
| | (1.793) | (1.564) | (1.461) | (1.344) |
| Allowed DI | 18.523*** | 12.010*** | 10.599*** | 8.238*** |
| | (2.674) | (2.242) | (2.027) | (1.869) |
| Dependent mean | 7.187 | 9.625 | 11.158 | 12.132 |
| **Panel D.** | | **Earnings ($1000)** | | |
| Judge leniency | -4.154* | -3.860* | -3.135 | -3.592* |
| | (2.318) | (2.252) | (2.377) | (2.137) |
| Allowed DI | -5.448* | -5.065* | -4.092 | -4.655 |
| | (2.969) | (2.970) | (3.105) | (2.843) |
| Dependent mean | 14.143 | 14.238 | 13.758 | 13.202 |

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

*Note:* This table displays the impact of judge leniency and DI allowance on DI participation (Panel A), labor force participation (Panel B), annual DI benefits (Panel C), and annual labor earnings (Panel D). All regressions include fully interacted year and department dummies, a linear term for average indexed earnings and dummy variables for month of appeal, county of residence, age at appeal, gender, foreign born, marital status, children below age 18, education, and a number of medical diagnoses. The control variables are measured prior to the appeal. Average indexed earnings is mean earnings for the last ten years prior to appeal.

Table 5: **Main Results II: Benefit substitution.**

| **Years after decision** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| **Panel A.** | | **Total transfers ($1000)** | | |
| Judge leniency | 8.414*** | 7.169*** | 6.625*** | 5.266*** |
| | (2.148) | (2.102) | (1.796) | (1.953) |
| Allowed DI | 11.035*** | 9.408*** | 8.648*** | 6.823*** |
| | (2.617) | (2.522) | (2.326) | (2.615) |
| Dependent mean | 19.519 | 20.047 | 20.510 | 21.024 |
| **Panel B.** | | **DI Benefits ($1000)** | | |
| Judge leniency | 14.123*** | 9.151*** | 8.120*** | 6.358*** |
| | (1.793) | (1.564) | (1.461) | (1.344) |
| Allowed DI | 18.523*** | 12.010*** | 10.599*** | 8.238*** |
| | (2.674) | (2.242) | (2.027) | (1.869) |
| Dependent mean | 7.187 | 9.625 | 11.158 | 12.132 |
| **Panel C.** | | **Other transfers ($1000)** | | |
| Judge leniency | -5.855** | -2.662 | -1.257 | -1.344 |
| | (2.374) | (1.857) | (1.601) | (1.898) |
| Allowed DI | -7.679** | -3.493 | -1.640 | -1.741 |
| | (3.176) | (2.507) | (2.074) | (2.408) |
| Dependent mean | 13.305 | 11.247 | 9.990 | 9.371 |
| **Panel D.** | | **Social assistance ($1000)** | | |
| Judge leniency | -1.497* | -1.159 | -1.409** | -0.648 |
| | (0.834) | (0.776) | (0.669) | (0.552) |
| Allowed DI | -1.963* | -1.522 | -1.839** | -0.840 |
| | (1.120) | (0.989) | (0.850) | (0.705) |
| Dependent mean | 2.865 | 2.193 | 1.800 | 1.486 |

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

*Note:* This table displays the impact of judge leniency and DI allowance on total transfers (Panel A), annual DI benefit payments (Panel B), other transfers except DI benefits (Panel C), and social assistance payments (Panel D). All regressions include fully interacted year and department dummies, a linear term for average indexed earnings and dummy variables for month of appeal, county of residence, age at appeal, gender, foreign born, marital status, children below age 18, education, and a number of medical diagnoses. The control variables are measured prior to the appeal. Average indexed earnings is mean earnings for the last ten years prior to appeal.

Table 6: **Main Results III: Household disposable income.**

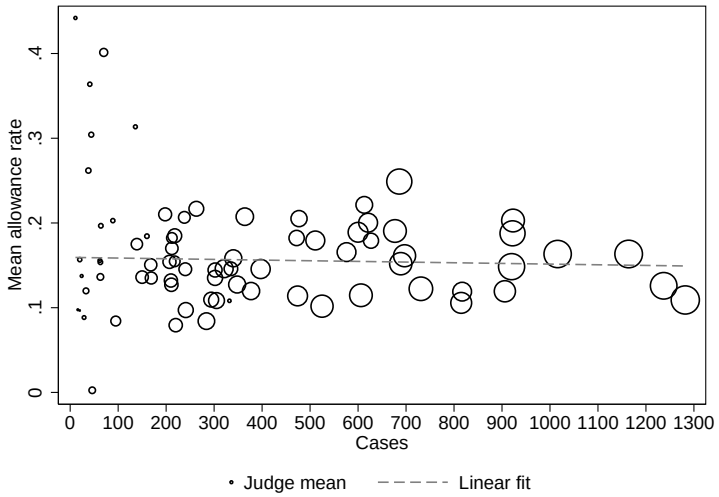
| **Years after decision** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| **Panel A.** | **Household earnings ($1000)** | | | |
| Judge leniency | -3 206 | -4.123* | -4.713** | -5.351** |
| | (2.033) | (2.093) | (1.951) | (2.176) |
| Allowed DI | -4 218 | -5.424* | -6.172** | -6.956** |
| | (2.721) | (2.805) | (2.635) | (3.028) |
| Dependent mean | 16.628 | 16.276 | 16.079 | 15.674 |
| **Panel B.** | **Household total transfers ($1000)** | | | |
| Judge leniency | 5.033*** | 6.615*** | 6.429*** | 6.170*** |
| | (1.314) | (1.118) | (1.204) | (1.757) |
| Allowed DI | 6.623*** | 8.703*** | 8.421*** | 8.022*** |
| | (1.746) | (1.541) | (1.841) | (2.615) |
| Dependent mean | 13.727 | 14.391 | 14.979 | 15.641 |
| **Panel C.** | **Household disposable income ($1000)** | | | |
| Judge leniency | 3 132 | 5.964*** | 3.142* | 3.757** |
| | (2.119) | (2.068) | (1.670) | (1.728) |
| Allowed DI | 4 122 | 7.847*** | 4.115* | 4.884** |
| | (2.807) | (2.667) | (2.109) | (2.397) |
| Dependent mean | 25.983 | 26.463 | 26.872 | 27.335 |

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

*Note:* This table displays the impact of judge leniency and DI allowance on household earnings (Panel A), household total transfers (Panel B), and household disposable income (Panel C). All regressions include fully interacted year and department dummies, a linear term for average indexed earnings and dummy variables for month of appeal, county of residence, age at appeal, gender, foreign born, marital status, children below age 18, education, and a number of medical diagnoses. The control variables are measured prior to the appeal. Average indexed earnings is mean earnings for the last ten years prior to appeal.

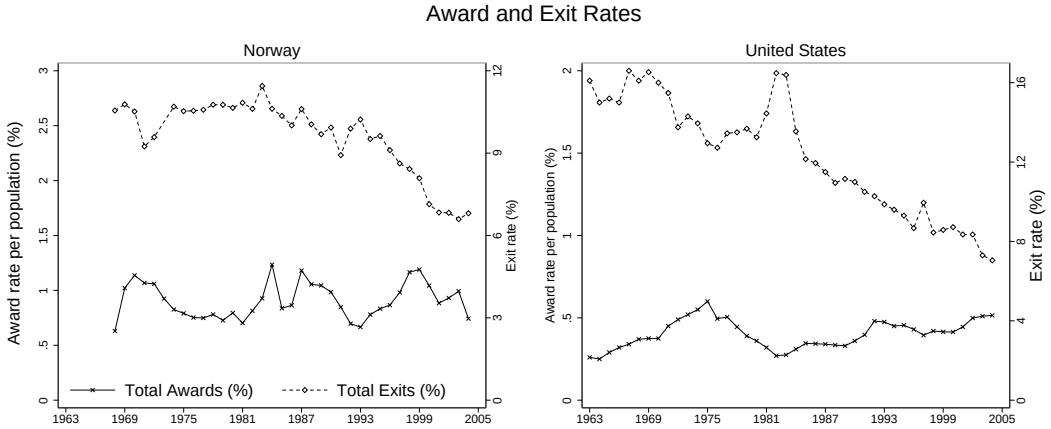
# Appendix

Figure A.1: **Judge Leniency versus Number of Cases Handled.**



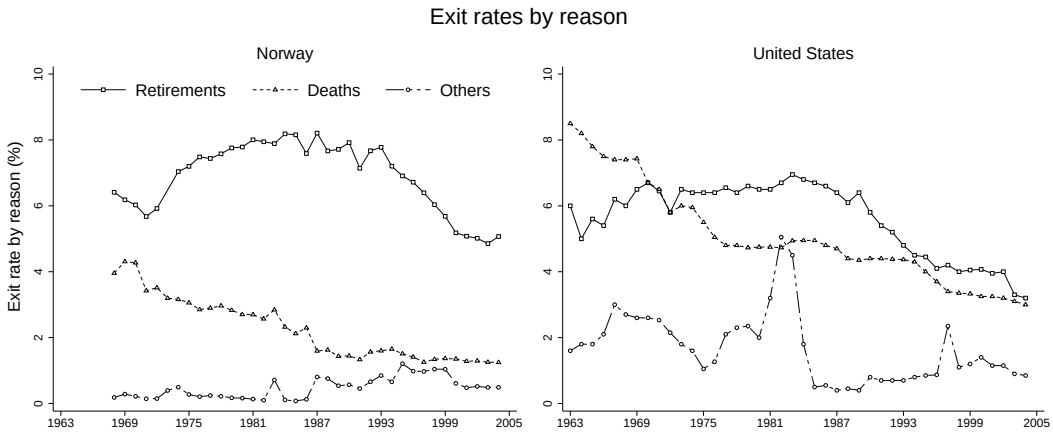

*Notes:* The figure plots a judge's allowance rate against the total number of cases he or she has handled. There are 76 different judges, and on average, each judge has handled a total of 375 cases. Allowance rates normalized by subtracting off year × department deviations from the overall mean. The sample is restricted to individuals appealing their first denied case during the period 1993-2005. Dot size is proportional to the number of cases a judge handles in the estimation sample (which is weakly smaller than the number of cases they have ever handled, as plotted on the x-axis).

Figure A.2:




*Notes:* The U.S. trends are based on Autor & Duggan (2006), while the Norwegian trends are collected from various issues of the SSA Supplement. The graphs show award rates in the insured population and exit rates from the DI program in both countries.

Figure A.3:




*Notes:* The U.S. trends are based on Autor & Duggan (2006), while the Norwegian trends are collected from various issues of the SSA Supplement. The graphs show exit rates because of death, retirement or other reasons (including eligibility-based exits).

Table A.1: **Characteristics of DI recipients in Norway and the U.S.**

| Characteristic | Norway DI Recipients | U.S. SSDI Recipients |
|---|---|---|
| Difficult to verify disorder | 59.2 % | 57.3 % |
| Age (at decision on initial application) | 52.2 | 49.1 |
| Prior earnings relative to the median | 71.0 % | 69.9 % |

*Notes:* The U.S. numbers are from Maestas *et al.* (2013), and the Norwegian numbers are drawn from the sample of DI applicants during the years 2000-2003. Difficult to verify disorder includes musculoskeletal and mental diagnoses. Prior earnings are measured 3-5 years before the application/appeal.

Table A.2: **Characteristics of DI Applicants and Appellants in Norway and the U.S.**

| | Norway | | U.S. | |
|---|---|---|---|---|
| Characteristic | Applicants | Appellants | Applicants | Appellants |
| Difficult to verify disorder | 60.9% | 69.7% | 58.5% | 62.2% |
| Age (at decision on initial application) | 51.1 | 47.1 | 47.1 | 46.1 |
| Prior earnings relative to the median | 66.5% | 50.4% | 60.5% | 56.3% |

*Notes:* This table displays the key characteristics of DI applicants and appellants discussed in Section 2. The U.S. numbers are from Maestas *et al.* (2013), and the Norwegian numbers are drawn from the sample of DI applicants during the years 2000-2003. Difficult to verify disorder includes musculoskeletal and mental diagnoses. Prior earnings are measured 3-5 years before the application/appeal.