Disability Benefits, Consumption Insurance, and Household Labor Supply*

David Autor† Andreas Ravndal Kostøl‡ Magne Mogstad§

September 1, 2015

Abstract: While a mature literature finds that Disability Insurance (DI) receipt discourages work, the welfare implications of these findings depend on two rarely studied economic quantities: the full cost of DI allowances to taxpayers, summing over DI transfer payments, benefit substitution to or from other transfer programs, and induced changes in tax receipts; and the value that individuals and families place on receiving benefits in the event of disability. We comprehensively assess these missing margins in the context of Norway's DI system, drawing on two strengths of the Norwegian environment. First, Norwegian register data allow us to characterize the household impacts and fiscal costs of disability receipt by linking employment, taxation, benefits receipt, and assets at the person and household level. Second, random assignment of DI applicants to Norwegian judges who differ systematically in their leniency allows us to recover the causal effects of DI allowance on individuals at the margin of program entry. Accounting for the total effect of DI allowances on both household labor supply and net payments across all public transfer programs substantially alters our picture of the consumption benefits and fiscal costs of disability receipt. While DI denial causes a significant drop in household income and consumption on average, it has little impact on income or consumption of married applicants; spousal earnings and benefit substitution entirely offset the loss in DI benefit payments. To develop the welfare implications of these findings, we estimate a structural model of household labor supply that translates employment decisions of both spouses into revealed preferences for leisure and consumption. We find that household valuation of receipt of DI benefits is considerably greater for single and unmarried individuals than for married couples, suggesting that it might be efficient to lower replacement rates or impose stricter screening on married applicants.

Keywords: disability insurance; consumption insurance; household labor supply; added worker
JEL codes: I38, J62, H53

---

*This research was supported by the U.S. Social Security Administration through grant #1 DRC12000002-03 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.The project also received financial support from the Norwegian Research Council. We are grateful to Richard Blundell, Raj Chetty, Amy Finkelstein, Kai Liu, Luigi Pistaferri, and Alessandra Voena for valuable input and guidance, to Bradley Setzler for excellent research assistance, and 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; 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., SSDI benefits receipt has risen from less than 1 percent to nearly 5 percent of the non-elderly adult population. In many European countries, the increases are even more striking, from 1 percent to 7 percent in the U.K and from 2 percent to almost 10 percent 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 primary cash benefit available after unemployment benefits run out and it has therefore become an increasingly important component of the social safety net.

To limit DI program growth, several countries have significantly tightened disability screening criteria, and many others are considering similar policies.[2] These enhanced gatekeeping policies can reduce the fiscal burden of disability insurance, both by lowering the DI caseload and, if rejected applicants return to work, by increasing tax revenue. At the same time, stricter screening may result in net welfare losses for individuals and families that value public disability insurance at more than its fiscal cost.[3] Assessing this tradeoff requires a comparison of the public costs and private benefits of DI awards for applicants at the margin of allowance versus denial, since it is their outcomes that would be changed by shifts in screening stringency. To implement this comparison, we need data on two economic quantities that are rarely measured: the economic value that individuals and families place on disability insurance; and the full cost of DI allowances to taxpayers, summing over DI transfer payments, benefit substitution to or from other transfer programs, and induced changes in tax receipts. Credibly estimating these quantities is typically hindered both by a lack of comprehensive linked data measuring the many outcomes, and by the difficulty of distinguishing the causal effects of DI receipts from the many unobserved factors that simultaneously determine disability status, earnings, tax payments and transfer receipts, and consumption.

This paper overcomes both the measurement and the identification challenge in the context of Norway's DI system to offer empirical evidence on the fiscal costs, consumption benefits and welfare implications of DI receipt. Our work draws on two strengths of the Norwegian environment. First, Norwegian register data allow us to characterize the household impacts and fiscal costs of disability receipt by linking employment, taxation, benefits receipt, and assets at the person and

[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 percent of GDP in the U.S. and 2.3 percent 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.

[3] In the U.S., all private disability insurance is provided through employer-based group policies. These policies 'wrap-around' the public SSDI system, so that most of the wage insurance risk and all of the medical cost risk is ultimately borne by the public program (Autor *et al.* , 2014b). There is not a strong standalone private market in disability insurance, likely because of adverse selection. In the Norwegian setting that we study, private disability insurance is rare.

household level. Our measure of fiscal costs includes virtually all forms of government cash transfers and revenues from (direct) taxes, accounting for changes in labor supply and substitution to other transfer programs. Our measures of household impacts of DI receipt include net government transfer payments from all sources, employment and earnings of DI applicants (both allowed and denied) and their spouses, and household consumption expenditure imputed from successive annual observations of income and wealth. Second, we address the threats to identification by exploiting the random assignment of DI applicants to Norwegian judges who differ systematically in their leniency. This approach recovers the causal effects of DI allowance on individuals at the margin of program entry. As a measure of judge leniency, we use the average allowance rate in all other cases a judge has handled. This leniency measure is highly predictive of judicial rulings in incumbent cases but uncorrelated with case characteristics.

Our first set of analyses estimating the causal effects of DI receipt on earnings, consumption, and fiscal costs yields three main findings. First, denying DI benefits to applicants on the margin of program entry induces an increase in annual earnings of approximately $6,600, which is about 40 percent of the annual DI transfer benefit denied. Second, DI denial lowers average household income and consumption by 15 and 16 percent—a reduction of approximately 60 cents for every dollar in net government spending averted—implying that DI receipt provides partial consumption smoothing. Third, DI denials have starkly different impacts on applicants according to marital status. Among single and unmarried (though possibly cohabiting) applicants, DI-induced changes in net government spending have large direct impacts on household income and consumption: each public dollar saved through DI denial reduces household income by nearly 90 cents. Conversely, DI denials do not decrease the household income or the consumption of married applicants. The reason is that household labor supply and benefit substitution entirely offset the loss in DI benefit payments. While provision of DI benefits does not increase consumption of married applicants, it does however impose considerable costs on other taxpayers through higher transfers and reduced payroll tax revenues. Thus, accounting for the total effect of DI allowances on household labor supply and net payments across all public transfer programs alters our picture of the consumption benefits and fiscal costs of disability receipt.

To develop the welfare implications of these findings, we estimate a structural model of household labor supply that translates employment decisions of both spouses into revealed preferences for leisure and consumption. The model allows for non-separable preferences between labor supply and consumption and the utility of leisure among spouses. Brought to the data, the model provides a good fit to the causal estimates of the impact of DI allowances on employment and total household income obtained non-structurally, and moreover, provides plausible parameter estimates for labor supply elasticities. We use the estimated model to compute the welfare benefits of DI receipt—by which we mean the cash equivalent value of receiving a DI allowance—and to perform counterfactual analyses that help us infer the extent to which spousal labor supply and reapplication attenuate the welfare loss from DI denials at the appeal stage. Among married couples, there is a relatively small positive welfare benefit of DI receipt, due to increased leisure of applicants and their spouses. By comparison, the welfare gains of single and unmarried applicants are large, and almost entirely due to increased disposable income. These results suggest that it might be efficient to lower replacement rates or impose stricter screening on married applicants. Of course, any policy that conditioned disability screening and benefits on marital status would have to account for likely policy-induced shifts in marriage formation and dissolution.

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, 2011; Liebman, 2015). While the mature literature on the impacts of disability benefits focuses primarily on the employment and earnings effects of DI allowance, little is known about the fiscal costs or consumption benefits.[4] Meyer & Mok (2013) and Kostol & Mogstad (2015) offer to our knowledge the only prior study that documents changes in income and consumption that follow changes in health and disability. Low & Pistaferri (2012) provide simulations from a life-cycle model to compare the insurance value and incentive costs of DI benefits. Our identification strategy, which uses judge assignments to isolate quasi-experimental variation in disability allowances, builds on three recent studies using U.S. data to estimate labor supply impacts of DI receipt. Exploiting quasi-experimental variation in DI allowances stemming from differences in disability examiner leniency, Maestas *et al.* (2013) and Autor *et al.* (2014a) find that DI receipt substantially reduces earnings and employment of applicants. French & Song (2013) pursue a similar strategy—exploiting variation in the leniency of appeal judges rather than initial examiners—and find comparable labor supply effects of DI receipt among appellants. Our study makes two contributions to this active literature. It combines quasi-experimental variation in judicial disability determinations with extensive register data on disability applicants and household members to provide novel evidence on the consumption benefits and fiscal costs of DI receipt.[5] Second, by integrating causal impact estimates along multiple dimension, the subsequent structural model estimation offers a welfare assessment of these findings.

Our paper also advances the study of household responses and consumption to income changes.[6] Most work in this literature assumes exogenous labor supply, focuses on a single earner, or imposes restrictions on the nature and type of insurance available to families. A notable exception is Blundell *et al.* (2012), who estimate a life cycle model with two earners making consumption and labor supply decisions.[7] Consistent with our findings, Blundell et al. find an important role for consumption insurance through household labor supply, while self-insurance through savings and borrowing matter less. In line with these results, Fadlon & Nielsen (2015) find that wives offset income losses

[4] This literature includes Parsons (1980), Bound (1989), Gruber (2000), Chen & van der Klaauw (2008), and Kostol & Mogstad (2014) as well as the methodologically related papers on DI discussed immediately below. See also Autor & Duggan (2003) and Borghans *et al.* (2014) for empirical evidence on the interaction between disability insurance and other transfer programs in the U.S. and Netherlands.

[5] Our analysis uses the same identification strategy as Dahl *et al.* (2014) though applied to a distinct question and set of outcomes.

[6] The literature is reviewed in Blundell *et al.* (2008), Meghir & Pistaferri (2011) and Blundell *et al.* (2012).

[7] A complementary exception is Finkelstein *et al.* (2015), who directly estimate the insurance value of Medicaid in-kind public health plan benefits using variation from a randomized controlled trial. Distinct from our focus, their work (a) abstracts from labor supply considerations since labor supply appears unaffected by Medicaid provision in their setting (Baicker *et al.* , 2014); and (b) estimates both the transfer and ex ante insurance values of public benefits provision, whereas we estimate only the first component.

following the death of a spouse through increased labor supply.

A related literature tests for the added worker effect, i.e., an increase in spousal labor supply induced by negative income shocks to the other spouse (Lundberg, 1985). Cullen & Gruber (2000) review this literature and highlight the difficulty in drawing credible inferences from observational data. One challenge is to locate a plausibly exogenous to the income of one spouse that does not directly affect the labor supply of the other spouse, thus overcoming the problem of reflection or simultaneity. Another difficulty is to control for correlated unobserved spousal heterogeneity in earnings capacity, health, and the taste for work, all of which might bias estimates of an added worker effect. A third challenge is to avoid or model correlated shocks across spouses. If, for example, a general economic downturn causes a negative income shock to a primary earner, his or her spouse's market wage will likely fall concurrently, thus biasing downward the estimated added worker effect. Our research design overcomes these challenges by identifying a plausibly exogenous income shock (DI denial) that directly affects one member of the household (the DI applicant), thereby providing a strongly confirmatory test of added worker effects in the DI context.

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 research design. In Section 3, we describe the data and sample restrictions. Section 4 assesses the relevance and validity of our instrument. Section 5 presents our findings on how the applicants respond to being denied versus allowed DI, and discusses the estimates of spousal responses to the allowance decision. Section 6 presents our findings on the fiscal costs and consumption benefits of disability receipt. Section 7 describes the structural model of household labor supply and explores welfare implication of disability receipt. The final section offers some concluding remarks.

2 Background

Following an institutional and statistical description of the Norwegian DI program, we document how the DI system generates quasi-random disability allowances for a subset of DI appellants (i.e., applicants who appeal their initial denial) and explain how our research design uses this variation to estimate the economic consequences of DI allowances.

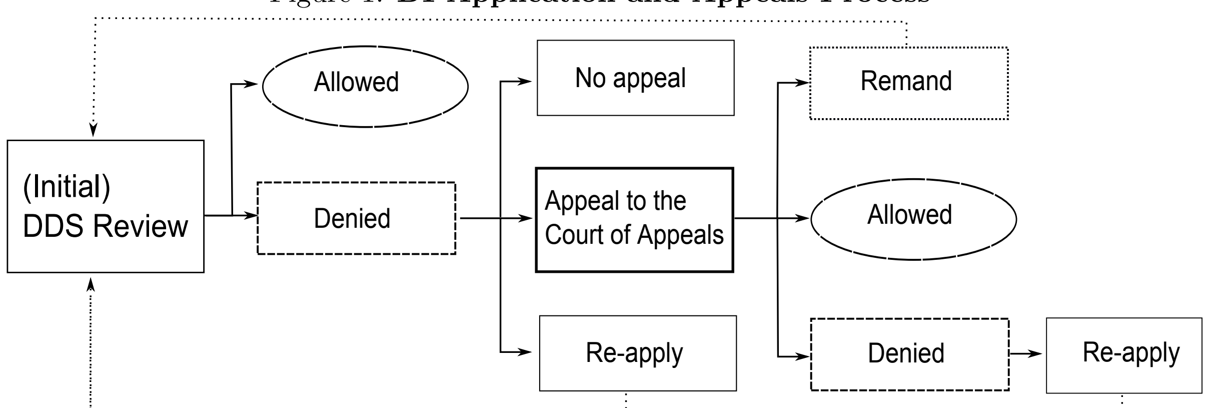
The Norwegian DI program

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 benefits schedule is progressive, so that low-wage workers replace a larger fraction of their earnings with DI benefits than do high-wage workers.

The disability determination process involves multiple steps, 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 the applicant's health status, age, education, work experience and the transferability of her skills. If the disability examiner concludes that the applicant cannot be expected to engage in any substantial gainful activity, a disability award is made. Approximately 75 percent of applicants are awarded benefits at the DDS review. About 30 percent of beneficiaries receive partial awards. Cases that are more difficult to assess—typically claims of mental illness and lower back pain—are often denied at this step.

Applicants who are denied at the DDS review may appeal within two months to the Court of Appeals, and about 25 percent of denied applicants do so. Appellants are assigned to Administrative Law Judges (ALJs), who either allow, deny, or remand (i.e., return to the DDS for reevaluation) their cases.[8] ALJs are required to apply the same criteria used in the initial determination process, but applicants may present new supporting information in writing. Approximately 15 percent of all appealed claims are allowed at the ALJ level. If the appeal is denied, the applicant can choose to start a new DI case by reapplying to the DDS Review stage.

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



Notes: This figure summarizes the description of the application and appeal process in the Norwegian DI system.

Assignment of DI cases to judges

All Norwegian disability appeals are heard in Oslo. Prior to 1997, there was only one hearing department; subsequently, there were four equally sized departments, all housed in the same building, and with no specialization across the four departments. Within each department, the assignment of cases to Administrative Law Judge is performed by a department head who does not have knowledge

[8] Average processing time at the DDS stage is six months, and average processing time at the appeal stage is four months. Remands are uncommon, accounting for only five percent of appeal outcomes. In our baseline analysis, we code remanded cases as rejected. As a robustness check, we recoded remanded cases as allowed or denied based on their eventual outcome after they were reconsidered by the DDS case worker. Results are comparable in either case.

of the content of cases. As stipulated in the rules set forth for the Administrative Law Court, case 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.[9]

Unusual among national disability systems, Norwegian judges are not specialized according to cases characteristic (medical condition, geographic region, or other attributes), and there is never any personal contact between judges and appellants (all evidence is presented in writing). Appellants have no option to seek a different judge from the one to whom they are assigned.

Verifying random assignment

Table 1 verifies that the hearing office assignment mechanism generates a distribution of cases across judges that is consistent with random assignment. There are 75 judges in our sample who have handled, on average, 375 cases each. We measure judge leniency as the average allowance rate in all other cases a judge has handled (including the judge's past and future cases that may fall entirely outside of our estimation sample). To purge any differences over time or across departments in the characteristics of appellants or the overall leniency rate of the DI system, we always control for fully interacted year by department dummies (the level at which randomization occurs.)

The first column of Table 1 uses a linear probability model to test whether case characteristics are predictive of case outcomes. As expected, demographic, work and health variables are highly predictive of whether an appealed case is allowed. Column 3 assesses whether these same case characteristics are predictive of the leniency of the judges to which cases are assigned. Reassuringly, we find no relationship. Jointly, these 16 variables explain less than 0.1 percent of the variation in the judge leniency measure (joint p-value of 0.49), and none is statistically significant at the 10 percent level.

A natural question is why some judges are more lenient than others. We have few detailed characteristics of judges to help illuminate this question, but we do know the number of cases that each judge has handled. We find that experienced judges appear to be slightly less lenient, but experience accounts for only a small fraction of the total variation in allowance rates across judges (see Appendix Figure A.1). While interesting for thinking about the design of the disability determination process, it is not critical to our analysis to know precisely why some judges are more lenient than others. What is critical is that appellants are randomly assigned to judges (as our data confirm), that some judges are systematically more lenient than others (as documented in Section 4.1), and that cases allowed by a strict judge would also be allowed by a lenient one (as the tests in Section 4.2 suggest).

[9] 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 work 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.

Instrument and 2SLS model

We use variation in DI allowance generated from the random assignment of appeal judges as an instrument to estimate the economic consequences of disability receipt. Because some judges are systematically more lenient than others, as we document below, random assignment of appellants to judges rise to exogenous variation in the probability an individual is allowed DI in the appeals process.

Our baseline instrumental variables (IV) model can be described by the following two-equation system:

$$A_i = \gamma Z_{ij} + X_i' \delta + \varepsilon_{ij} \quad (1)$$

$$Y_{it} = \beta_t A_i + X_i' \theta_t + \eta_{it}. \quad (2)$$

Here, A_i is an indicator variable equal to 1 if appellant i is allowed DI at the appeal, Z_{ij} denotes the leniency measure for judge j to whom appellant i was assigned, X_i is a a vector of control variables that includes a full set of year-of-appeal by department dummies, and Y_{it} is a dependent variable of interest that is measured for individual i at some point t after his allowance decision (e.g. earnings three years after the decision).

The target of our estimation is the average of β_t among individuals who are allowed DI at the appeal because they were assigned to a lenient judge. To estimate this parameter, our baseline specification uses 2SLS with first and second stage equations given by (1) and (2).[10] The endogenous variable in our estimation is an indicator for whether an appellant was allowed DI *at appeal*, rather than whether or not the appellant is currently receiving DI when outcome Y_{it} is observed. This specification alleviates concerns about the exclusion restriction: 2SLS estimates of β_t capture the causal effects of the *initial* judicial disability determination, which may operate through a number of channels, including participation in DI, subsequent reapplications to the DI program following denial, or other behavioral changes resulting from the initial outcome at appeal. We can also estimate the reduced form effect of judge leniency on appellant outcomes by directly regressing Y on Z and X.

[10] Our estimates of β_t do not change appreciably if we instead follow Doyle (2008) in using limited information maximum likelihood (LIML) or 2SLS with judge fixed effects as instruments.

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)	0.0044***	(0.0003)	0.0001	(0.0001)
Number of persons in household	-0.0159***	(0.0018)	-0.0004	(0.0003)
Average indexed earnings	0.0011***	(0.0002)	0.0000	(0.0000)
Female	0.0194***	(0.0057)	0.0008	(0.0012)
Married	0.0145**	(0.0066)	0.0005	(0.0012)
Foreign born	-0.0443***	(0.0086)	-0.0003	(0.0015)
Less than high school degree	-0.0230***	(0.0061)	-0.0005	(0.0008)
High school degree	0.0196***	(0.0061)	0.0001	(0.0007)
Any college	0.0114	(0.0115)	0.0009	(0.0014)
Children below age 18	-0.0573***	(0.0051)	-0.0009	(0.0008)
Musculoskeletal disorders	-0.0171***	(0.0059)	0.0005	(0.0017)
Mental disorders	0.0089	(0.0075)	-0.0002	(0.0024)
Circulatory system	0.0234	(0.0158)	0.0000	(0.0023)
Respiratory system	-0.0197	(0.0151)	-0.0021	(0.0021)
Neurological system	0.0458**	(0.0206)	0.0011	(0.0021)
Endocrine diseases	0.0417***	(0.0174)	-0.0029	(0.0031)
F-statistic for joint significance	28.20		0.97	
[p-value]	[.001]		[.49]	
Observations	14,077		14,077	

***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 1994-2005 (see Section 3 for further details). There are 75 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 appellant characteristics. Each regression controls for fully interacted year of appeal and department dummies. 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.

3 Data and Background

3.1 Data and Sample Restrictions

Our analysis draws on multiple administrative data sources that are linked using unique person-level identifiers. 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. These data include 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. These person-level records are linked to hearing office records on all DI appeals during 1989 through 2011, including dates of appeal and decision, outcomes for each appeal, and unique identifiers for both judges and appellants.

To capture complete information on DI applicants' earnings, assets, and implicitly consumption, we merge the social security data with longitudinal administrative registers provided by Statistics Norway covering every Norwegian resident from 1967 to 2010. These register data enumerate individual demographic information (including sex, age, and education) and, since 1993, all sources of annual income, including earnings, self employment income, capital income, and cash transfers, as well as most types of assets holdings and liabilities, such as as real estate, financial portfolio, and debt. Income data are reported in annual amounts, while the values of assets holdings and liabilities are measured as of the last day of each year.

The Norwegian data have several advantages over register data collected by many other countries. Because most components of income and wealth are third-party reported (e.g. by employers, banks and financial intermediaries), the coverage and reliability are rated as exceptional by international quality assessments (see e.g Atkinson *et al.* 1995). Because in Norway, most register data are a matter of public record, there is no attrition from the original sample due to non-response or non-consent. The income and wealth data pertain to all Norwegian residents, and are therefore not limited to those employed in jobs covered by social security, individuals who respond to wealth surveys, or households that file estate tax returns. Measures of incomes and wealth are recorded without any top or bottom coding.[11] Finally, unique identifiers allow us to match spouses to one another and parents to children, thereby constructing measures of per capita household income and consumption.

Our empirical analysis considers individuals who appeal an initially denied DI claim.[12] To observe individuals for at least four years after the appeal decision, our estimation sample consists of individuals whose appeal decision was made during the period 1994-2005. Following Maestas *et al.* (2013) and French & Song (2013), our baseline estimation excludes observations for which the assigned appeal judge has handled fewer than 10 cases during the 1989 through 2011 period.[13] This restriction reduces noise in our instrument stemming from sampling variation. To circumvent the issue of older appellants substituting between DI and early retirement, we also exclude appellants above age 62 years at the time of appeal,

In Table 2, we document characteristics of the sample of individuals who apply for DI and the subsample who appeal an initially denied DI claim (our baseline sample). Relative to the full sample of initial applicants, those who appeal are more likely to be female, are less educated, are more likely to be foreign born, and have lower prior earnings and liquid assets. Sixty-three percent of applicants

[11] Some individuals are reported with negative income components (e.g. negative cash transfers). In these cases, we truncate the income components at zero. We also top-code a handful of observations with extremely large income components. The results do not change appreciably if we retain these outliers.

[12] 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.

[13] Including these judges does not change the estimates appreciably, and neither does excluding judges who handle fewer than 50 cases.

claim mental or musculoskeletal disorders, whereas this figure is 70 percent for appellants.

Table 2: **Descriptive Statistics of Applicants and Appellants**

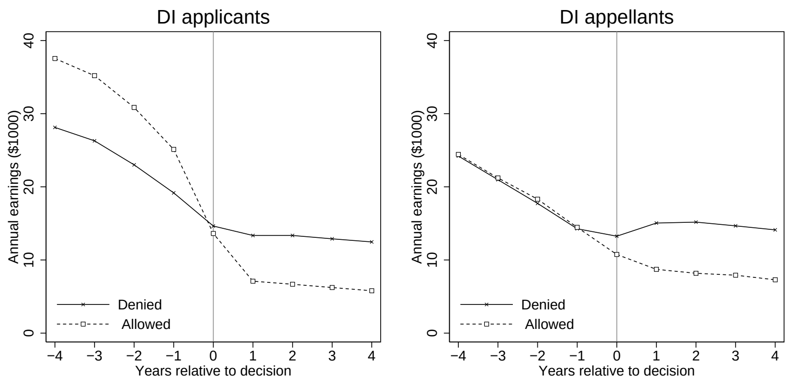
	DI applicants		**DI appellants**	
Characteristic	Mean	Std. Dev.	Mean	Std. Dev.
Age (at the time of decision)	48.46	[10.04]	46.61	[9.30]
Number of persons in household	3.22	[2.02]	2.86	[1.60]
Average indexed earnings	33.08	[23.63]	20.93	[18.19]
Female	0.56	[0.50]	0.63	[0.48]
Married	0.57	[0.50]	0.57	[0.49]
Foreign born	0.08	[0.27]	0.18	[0.38]
Less than high school degree	0.43	[0.50]	0.50	[0.50]
High school degree	0.42	[0.49]	0.39	[0.49]
Any college	0.13	[0.34]	0.11	[0.31]
Children below age 18	0.31	[0.46]	0.43	[0.50]
Musculoskeletal disorders	0.37	[0.48]	0.44	[0.50]
Mental disorders	0.26	[0.44]	0.26	[0.44]
Circulatory system	0.08	[0.27]	0.04	[0.19]
Respiratory system	0.03	[0.17]	0.03	[0.16]
Neurological system	0.06	[0.23]	0.04	[0.19]
Endocrine diseases	0.02	[0.14]	0.04	[0.20]
Liquid assets ($1,000)	39.71	[70.43]	10.63	[21.18]
DI allowed	0.78	[0.42]	0.13	[0.33]
Observations	239,903		14,077	

Standard deviations [in square brackets]

Notes: This table displays descriptive statistics for applicants and appellants. The applicant sample consists of all claims made during the period 1992-2003 by individuals who are at most 61 years of age. The appellant sample filed an appeal during the period 1994-2005 (see Section 3 for further details). 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.

Figure 2 summarizes the earnings trajectories of DI applicants and appellants before and after their DI determinations, focusing on years $t-4$ through years $t+4$ surrounding the initial DI determination (panel A) and the year of the initial appeal decision (panel B). Consistent with U.S. data, panel A shows that DI applicants who are allowed at the initial determination have on average substantially higher prior earnings than those who are denied (von Wachter *et al.* , 2011; Maestas *et al.* , 2013). This pattern may reflect the fact that workers with high prior earnings who seek DI benefits face severe impairments that necessitate a cessation of employment whereas applicants with low prior earnings may be more motivated by financial considerations. By contrast, there is no discernible difference in the pre-determination earnings of DI appellants who are allowed versus those who are denied at appeal.[14] Earnings between these groups diverge immediately after the appeal decision, however, and this gap does not close over the subsequent four post-decision years.

Figure 2: **Earnings Trajectories of Allowed and Denied DI Applicants and Appellants**



Notes: This figure displays mean real earnings for denied and allowed DI applicants (left-hand panel) and DI appellants (right-hand panel) in the nine years surrounding the initial DI determination (left-hand panel) and the initial outcome at appeal (right-hand panel). The applicant sample consists of all claims made during the period 1992-2003 by individuals who are at most 61 years of age. The appellant sample filed an appeal during the period 1994-2005 (see Section 3 for further details). Nominal values are deflated to 2005 and represented in US dollars using the average exchange rate NOK/$ = 6.

Table 3 shows OLS estimates of the raw differences in outcomes between allowed and denied appellants after the allowance decision. For brevity, we focus on averages of the outcomes over the first five years (inclusive of the year of the decision). Since individuals who are denied at appeal differ in many important ways from those who are allowed (see e.g. Column 1 of Table 1), these OLS estimates should be viewed as entirely descriptive.

Over the five years following appeal, individuals who are allowed DI benefits receive an additional $9,580 in annual DI benefits relative to those denied, and earn $5,850 less per year in income earnings. Among married appellants, spousal earnings in the five years following appeal are on average $1,950 higher among spouses of denied than allowed appellants. The last two rows of the table contrast household disposable income and consumption (per household member) among allowed versus denied appellants. On average, allowed appellants have an additional $3,400 ($3,020) per year in income (consumption), which amounts to 13 (11) percent of the sample mean of income (consumption).

[14]It is not necessary for our identification strategy that allowed and denied appellants are comparable, since random assignment of appellants to judges who differ in their leniency generates variation in appeal outcomes among a subset (complier) group of appellants that is independent of potential outcomes.

Table 3: **Simple OLS Regression of Appellant Outcomes on the Allowance Decision**

Dependent variable (5-year average, \$1,000)	**Allowed at Appeal** Coeff.	Std. Err.	Number of obs.	Dependent mean
Appellant:				
DI benefits	9.58***	(0.18)	14,077	6.55
Earnings	-5.85***	(0.32)	14,077	13.63
Spouse earnings (married appellants only)	-1.95*	(1.07)	8,349	39.66
Household:				
Disposable income (per capita)	3.40***	(0.26)	14,077	26.20
Consumption (per capita)	3.02***	(0.36)	14,077	28.70

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are robust to heteroskedasticity.

Notes: This table displays the OLS estimates of the differences in outcomes between allowed and denied appellants after the allowance decision. The dependent variables are measured as average outcomes over the first five post-determination years (including the year of decision). Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). When estimating the effect on spousal earnings, we restrict the sample to married appellants. Nominal values are deflated to 2005 and represented in US dollars using the average exchange rate NOK/\$ = 6.

3.2 Measuring Consumption

A key challenge in estimating the consumption benefits of disability insurance is the lack of reliable longitudinal data on consumption expenditures. One approach to measuring expenditures is to use survey data, but expenditure surveys typical have small sample sizes and face significant measurement issues (for a discussion, see Pistaferri, 2015). A second option is to create measures of consumption from longitudinal data on income and wealth. Browning & Leth-Petersen (2003) investigate the quality of such measures of consumption. Brinch *et al.* (2015) perform a similar exercise using highly detailed Norwegian data on income and assets. Their analysis shows that the measures of consumption derived from income and assets conform well to those reported in family expenditure surveys and to the aggregates from national accounts. We use their measures here.

Following Brinch *et al.* (2015), let the household of individual i begin year t with a portfolio (vector) of assets $\{V_{ikt-1}\}$, where V_{ikt-1} is the level of asset k at the end of period $t-1$. These assets are held throughout the year, with each asset k earning a return of r_{kt} . The household also receives income E_{it} (including market income and transfers) and pays taxes τ_{it}. At the end of the year, the household sells the assets V_{ikt-1} at prices p_{kt} and buys a new portfolio V_{ikt} at the same prices. The following identity links current consumption and asset holdings to prior asset holdings, asset returns, and contemporaneous income net of taxes,

$$C_{it} + \sum_k p_{kt} V_{ikt} = \left(E_{it} + \sum_k r_{kt} V_{ikt-1} - \tau_{it} \right) + \sum_k p_{kt} V_{ikt-1}, \tag{3}$$

where C_{it} denotes the household's consumption expenditure throughout the year. We can re-arrange (3) so that consumption equals disposable income minus the change in wealth plus capital gains:

$$C_{it} = \left(E_{it} + \sum_k r_{kt} V_{ikt-1} - \tau_{it} \right) - \sum_k (W_{ikt} - W_{ikt-1}) + \sum_k (p_{kt} - p_{kt-1}) V_{ikt-1} \tag{4}$$

where $W_{ikt} = p_{kt} V_{ikt}$ and the final term on the right hand side is the capital gains on the portfolio held at the beginning of the year.

If all the components on the right hand side of equation (4) were observed, we could construct a measure of household consumption expenditure directly. In reality, most but not all of these components are directly available from our register data. We have information for each individual and household on virtually all sources of income as well as most types of assets and durables (such as deposits, securities, liabilities, pension plans, real estate, vehicles) from 1993 forward. A few measurement issues remain, which we resolve following Brinch *et al.* (2015). Because the value of owner-occupied housing services is not observed, we use national accounts data on the aggregate value of housing services to assign to each homeowner an individual value according to the expected market value of their home as a share of total expected market value of all primary residences.[15] Because we do not observe the individual price change for every asset that a household holds, we impute the missing data from supplementary information on asset-specific price changes and rates of depreciation at a disaggregated level.

Appendix Figure A.4 compares the distributions of total expenditure from the Norwegian family expenditure surveys to the distributions of our measures of consumption derived from data on income and assets. Reassuringly, the two measures of consumption have similar distributions. Over time, however, the expenditure survey suggests lower levels of consumption as compared to our measures of consumption. A likely reason for the discrepancy is the problem of growing under-reporting in expenditure surveys. Pistaferri (2015) show that there is more severe understatement of spending among rich than non-rich households, and that survey participation among high income households has declined over time. This gives rise to a growing discrepancy between average consumption as measured in expenditure surveys (such as CEX in the United States) and national accounts aggregates. By comparison, our consumption measure follows more closely the time trend in average household consumption expenditure that we observe in the Norwegian national account data (see Appendix Figure A.5).

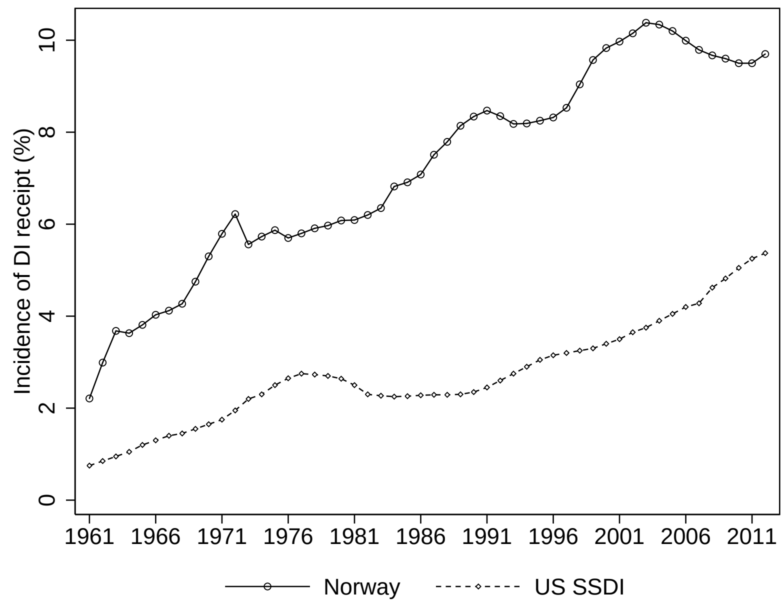
3.3 Institutional Background

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 prevalence of receipt of DI benefits is lower in

[15] National account data are derived from imputed rents for owners based on the market rents for observationally equivalent housing in the same area.

the U.S. than in Norway, as shown in Figure 3, while the time trends are similar.[16] From 1961 to 2012, DI prevalence increased from 2.2 to 9.7 percent in Norway and from 0.8 to 5.0 percent in the U.S. While Norway's prevalence has leveled off at about 10 percent in recent years, the U.S. DI rate continued to rise steeply through 2013, after which time growth slowed substantially (OASDI Trustees Report, 2014).

Figure 3: **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).

In both countries, the expansion of the DI rolls in recent decades appears to be driven in significant part by changes in disability screening criteria, which led to a steep rise in the share of DI recipients suffering from difficult-to-verify disorders such as mental illness and musculoskeletal disease.[17] 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. DI exit rates in both countries have consequently 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). In

[16] 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.

[17] See Autor & Duggan (2006) and Liebman (2015) for discussions 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.

addition, the aging of the Baby Boom cohorts into their peak (near-elderly) disability age brackets has contributed substantially to the expansion of the DI rolls since the mid-1990s (Liebman, 2015)

There are a few noteworthy differences between the U.S. and Norwegian DI programs. DI recipients in Norway tend to be older and have slightly higher earnings prior to a disability award (see Appendix Table A.1). A second difference is that the appeal process plays a more important role in the U.S. than in Norway. While 48 percent of initially rejected applicants appeal in the U.S. (French & Song, 2013), only 25 percent of those rejected appeal in Norway. Success rates at appeal are also considerably higher in the U.S. than Norway.

Despite these differences, there are key similarities between the participant populations. Almost 60 percent of DI recipients in both countries suffer from difficult-to-verify mental and musculoskeletal disorders (see Appendix Table A.1). And in both countries, appellants have a weaker prior connection to the labor market and are more likely to suffer from difficult-to-verify disorders than are applicants (see Appendix Table A.2).

4 Assessment of the Instrument

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

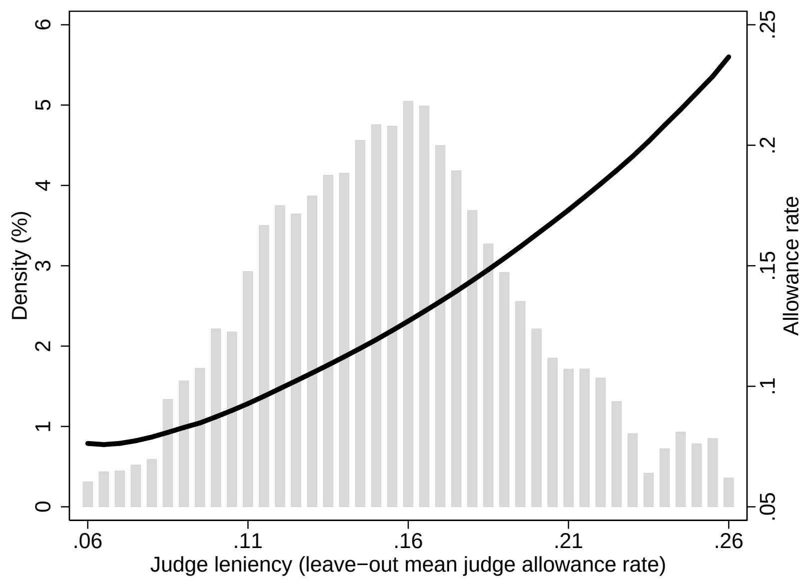
4.1 Instrument Relevance

Figure 4 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 (controlling for fully interacted year and department dummies). The measure of judge leniency is the average judge allowance rate in all other cases a judge has handled (including the judge's past and future cases that may fall entirely outside of our estimation sample). The mean of the leniency variable is 0.15 with a standard deviation of 0.05. The histogram reveals a wide spread in judge leniency, with a judge at the 90th percentile allowing approximately 18 percent of cases as compared to approximately 8 percent for a judge at the 10th percentile.

Figure 4 also graphs the relationship between judge leniency and the appellant's allowance rate (controlling for fully interacted year and department dummies). 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 that an individual appellant's case is allowed.

Table 4 presents estimates of our first equation for the relationship between judge leniency and DI allowance rates at appeal (1). 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 the judge leniency measure. The four columns of the table correspond to years one through four following appeal. These columns are identical except for the very modest impact of sample attrition (less than three percent over four years) stemming from death or emigration of appellants.[18] The point estimate of approximately 0.82 is essentially identical across columns, indicating that attrition exerts a negligible impact on the first stage relationship. This estimate indicates that, all else equal, being assigned to a judge with a 10 percentage point higher overall allowance rate increases the probability of receiving an allowance by 8.2 percentage points.

Figure 4: **Effect of Judge Leniency on DI Allowance**



Notes: This figure displays the effect of judge leniency on DI allowance, conditional on fully interacted year and department dummies. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 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 percent excluded from the graph).

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 4 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

[18]One might nevertheless be concerned that the instrument affects mortality or emigration rate. The first column of Appendix Table A.4 indicates that this is not the case.

Table 4: **First Stage: Judge Leniency and DI Allowance**

	Years after decision			
	1	**2**	**3**	**4**
Panel A.	**No covariates**			
Judge leniency	0.817***	0.819***	0.821***	0.828***
	(0.082)	(0.084)	(0.083)	(0.083)
Panel B.	**With individual covariates**			
Judge leniency	0.787***	0.786***	0.789***	0.794***
	(0.079)	(0.082)	(0.080)	(0.080)
Panel C.	**With judge characteristics**			
Judge leniency	0.803***	0.805***	0.810***	0.818***
	(0.075)	(0.076)	(0.075)	(0.076)
Dependent mean	0.13	0.13	0.13	0.13
Observations	13,957	13,826	13,694	13,593

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

Note: This table displays the first stage coefficients of equation 1. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 different judges. 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 comprising fully interacted year and department dummies dummy variables for month of appeal, county of residence, age at appeal, household size, gender, foreign born, marital status, children below age 18, education, and a number of medical diagnoses. In Panel C, we include judge leave-out-mean processing time.

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 impact of judge assignments on subsequent outcomes, our interpretation of the IV estimates as measuring the causal effect of DI allowances on appellant outcomes requires two additional assumptions. The first is that judge leniency affects appellant outcomes of interest only through its impact on the appellant's allowance decision, and not directly in any other way. This exclusion restriction appears particularly likely to hold in Norway because all appeals are presented in writing, so there is never any personal contact between judges and appellants; individuals (and their families) observe only judges' allowance or denial decisions. One potential violation of the exclusion restriction could occur if appeals processing times differed systematically with judge leniency and, moreover, exerted an independent effect on appellant outcomes (as explored in Autor *et al.* 2014a). To test this possibility, we calculated each judge's average processing time based on the residual average processing time in his or her other cases, controlling for a fully interacted set of time and department dummies. Panel C of Table 4 shows that the first stage estimates do not change appreciably when controlling for judge processing time.

The second additional assumption needed for a causal interpretation of the IV estimates is the monotonicity of the judge leniency instrument.[19] In our setting, this monotonicity assumption requires that cases allowed by a strict judge would also be allowed by a more lenient judge, and similarly that cases denied by a lenient judge would also be 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 (see Appendix Table A.3).

A second implication of monotonicity is that judges who are stricter towards one group of appellants (e.g., young appellants, those with mental disorders) are also relatively strict towards other applicants outside of this group (e.g., older appellants, those with musculoskeletal disorders). To test this implication, we categorize all cases a judge has handled into mutually exclusive subgroups based on the observable case characteristics used in Appendix Table A.3: marital status, disorder, gender, age, education, and liquid assets. We calculate each judge's leave-out-mean allowance rates in all cases within these subgroups, and then re-estimate the first stage equation (1) for each subgroup, using each judge's leniency measure for cases outside of that subgroup (e.g., the effect of judge leniency constructed from cases with younger appellants on the allowance rates of older appellants). In all cases, the first stage estimates are positive and sizable, suggesting that judges who are relatively strict for one subset of cases also are strict for other types of cases.[20]

5 Behavioral Responses to DI Allowance

5.1 Labor Earnings and DI Benefits

Table 5 reports 2SLS estimates of equations (1) and (2), with labor earnings, DI participation, and DI benefits as dependent variables. As in Table 4, we separately estimate the effects for each of the four years after the appeal decision. All specifications control for observable case characteristics and include a full set of year by department dummies.

The first two panels consider the impact of being allowed at the appeal on DI participation and benefits payments over the first four years following appeal. Column 1 of panel A reports a 2SLS point estimate of 0.99, indicating that allowances induced by judge leniency increase DI receipt almost one-for-one in the first year following appeal.[21] The declining pattern of coefficients across columns in panel A reveals that a substantial fraction of individuals who are initially denied DI reapply and are ultimately allowed.[22] Over the the first four years following appeal, the causal effect of being allowed at the appeal on subsequent DI receipt falls by approximately half, from 0.99

[19] If the treatment effect of the disability determination is constant across appellants, the monotonicity assumption is unnecessary.

[20] A table with this (large) set of estimates is available from the authors.

[21] Note that 0.991 = 0.780/0.787, where 0.787 is the corresponding first stage coefficient from Table 4, panel B column 1.

[22] Although this pattern could also be consistent with successful appellants exiting the DI program in years two through four, such exits rarely occur.

to 0.47. Thus, four years after the court decision, an allowance at the appeal increases the probability of current DI participation by 47 percentage points. Panel B displays analogous estimates for DI benefit payments. Receiving a DI allowance at appeal leads to a large increase in benefit payments relative to the alternative outcome, equal to $15,600 in the first year. But this contrast declines over time due to reapplication, reaching $8,900 in year four.

Panel C of Table 5 displays comparable estimates for annual labor earnings. We estimate sizable negative causal effects of DI allowance on labor earnings. Receiving a DI allowance at the appeal reduces annual earnings by approximately $6,600 in the first year after appeal, equal to approximately forty percent of the annual DI transfer benefit received. In contrast to the pattern over time for DI participation and benefits, the causal effect of DI allowances on appellant earnings declines only modestly over the four years following the appeal.

Table 5: **Effect of DI Allowance on Labor Earnings and DI Benefits of the Appellant**

	Years after decision			
	1	**2**	**3**	**4**
Panel A.	**DI participation**			
Allowed DI	0.991***	0.726***	0.647***	0.471***
	(0.071)	(0.103)	(0.099)	(0.085)
Dependent mean	0.305	0.432	0.519	0.577
Panel B.	**DI benefits ($1000)**			
Allowed DI	15.606***	14.037***	10.565***	8.915***
	(1.184)	(1.734)	(1.701)	(1.650)
Dependent mean	4.427	7.291	9.506	10.941
Panel C.	**Earnings ($1000)**			
Allowed DI	-6.590**	-5.767*	-5.354*	-5.489*
	(2.803)	(2.957)	(3.052)	(2.839)
Dependent mean	14.246	14.289	13.811	13.250
Observations	13,958	13,827	13,695	13,594

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

Note: This table displays the impact of being allowed at the appeal on DI participation (Panel A), annual DI benefits (Panel B), and annual earnings (Panel C) of the appellant. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 different judges. All regressions include fully interacted year and department dummies dummy variables for month of appeal, county of residence, age at appeal, household size, 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.

The IV estimates can be interpreted as local average treatment effects (LATE) for appellants whose DI decisions are causally affected by the instrument, meaning they could have received a different allowance decision had their case been assigned to a different judge. We can use these estimates to characterize how much compliers would have received in DI benefits and earned in labor income in the subsequent four years had they received a different initial decision at appeal. As shown in Dahl *et al.* (2014), we can recover these potential outcomes for compliers 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 A.6, we implement these calculations to decompose the LATE into the potential outcome of compliers if denied or, alternatively, if allowed at appeal. Relative to the regression estimates in Table 5, the Figure A.6 plots depict *levels* of potential outcome for compliers when allowed and when denied, rather than simply the contrast between the potential outcomes in the two states. We find that the labor earnings of compliers when denied change little over time, even though many reapply for, and eventually receive, DI. This pattern suggests that among the population of denied compliers, a small but non-negligible set works persistently following denial, while a larger group works little and persists in seeking and, in many cases, obtaining DI benefits.

5.2 Benefit Substitution

DI is one of several transfer programs available to Norwegians, and those denied DI may potentially substitute towards these other programs. Conversely, once allowed DI, disabled beneficiaries may seek additional transfer benefits. Table 6 reports 2SLS estimates for total transfers (DI benefits plus all other cash transfers) and other cash transfers excluding DI benefits. For comparison, we also repeat the estimated effects on DI benefit payments.

The results in Table 6 point to the importance of accounting for benefit substitution when considering the consequences of disability allowances, from both the perspective of public finances and household income and welfare. Table 6 demonstrate that the net impact of a DI allowance on total transfers received is substantially smaller than its gross impact. For each dollar in DI benefits paid as the result of a DI allowance, approximately 30 cents in expenditure from other transfer programs is averted.

Table 6: **Effect of DI Allowance on Transfer Payments, in Total and by Program**

	Years after decision			
	1	**2**	**3**	**4**
Panel A.	**Total transfers ($1000)**			
Allowed DI	10.508***	9.202***	8.380***	6.604**
	(2.583)	(2.612)	(2.364)	(2.645)
Dependent mean	19.569	20.070	20.543	21.054
Panel B.	**DI Benefits ($1000)**			
Allowed DI	15.606***	14.037***	10.565***	8.915***
	(1.184)	(1.734)	(1.701)	(1.650)
Dependent mean	4.427	7.291	9.506	10.941
Panel C.	**Other transfers ($1000)**			
Allowed DI	-5.015*	-4.794*	-2.348	-2.049
	(2.841)	(2.580)	(2.048)	(2.533)
Dependent mean	15.180	12.832	11.096	10.176

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

Note: This table displays the impact of being allowed at the appeal on total transfers (Panel A), annual DI benefit payments (Panel B), and other transfers except DI benefits (Panel C). Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 different judges. All regressions include fully interacted year and department dummies dummy variables for month of appeal, county of residence, age at appeal, household size, 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.

In Appendix Figure A.7, we decompose the LATEs for benefit receipt into potential outcomes for compliers when allowed and when denied. If compliers are awarded DI benefits, we see a sizable fall in their payments from non-DI transfer programs. By contrast, non-DI transfer payments change little in the year following appeal when compliers are denied DI. These transfers do, however, tail off over the next several years as many of those denied at appeal successfully reapply for DI benefits.

5.3 Spousal Responses

Table 7 extends our inquiry to consider household level responses to DI allowance. We focus exclusively on married households in this analysis since our data do not allow us to determine whether non-married appellants are single or cohabiting.[23]

As a baseline for comparison, panels A and B report 2SLS estimates of the impact of DI allowances on labor earnings and total transfer payments of married appellants. While the smaller sample size available for these estimates reduces precision, the point estimates suggest that DI denial has a similar, perhaps modestly smaller, impact on labor earnings and total (individual) transfer payments of married appellants as compared to the full appellant population.

[23]We test for endogenous selection into marital status according to disability application outcome in Appendix Table A.4 column 2. We find that the judge leniency instrument does not affect the likelihood of marital dissolution.

Panels C and D next consider the impact of DI appellant outcomes on potential compensatory behaviors among appellants' spouses. The 2SLS estimates show that even in the Norwegian context with considerable scope for benefit substitution, the appellants' spouses respond strongly to the outcomes of disability determinations. Spousal earnings of denied appellants rises by approximately \$6,000 in the year following denial (relative to spouses of allowed appellants), and by a further \$10,000 to \$12,000 in years two through four following denial (panel C). Spousal labor supply therefore substantially insures household income against the adverse shock of DI denial. Panel D reveals, however, that spouses face a high effective tax on these additional earnings; reductions in cash transfers to the spouse offset up to 50 percent of the increase in spousal earnings.

These estimates are consistent with the possibility that either spouses of allowed appellants reduce their labor supply, or that spouses of denied appellants increase their labor supply (or potentially both). The latter possibility would suggest that DI denials induce an added worker effect among spouses. The former possibility would instead suggest that DI allowance induces a decline in labor supply among spouses due to leisure complementarities or income effects. We explore which interpretation is consistent with the data by decomposing the causal effects estimates in Table 7 into potential outcomes of spouses of compliers when these appellants are denied or, counterfactually, allowed benefits. This decomposition, found in Appendix Figure A.8, indicates that the behavioral response found in Table 7 stems almost entirely from spousal responses to DI denials: when appellants are denied, their spouses respond strongly by increasing earnings in the years following denial; conversely, if these appellants are instead denied, their spouses exhibit little earnings adjustment. Thus, the effect of DI denial induces a powerful added worker effect on spouses.

We have also explored the heterogeneity of the added worker effect among households according to the spouse's sex, education, and prior earnings. We find suggestive evidence that the labor supply of female spouses is more responsive to DI denial than is the labor supply of male spouses. Due to small sizes, however, these contrasts are not significant at conventional levels.

Table 7: **Effect of DI Allowance on Spousal Earnings and Transfer Payments**

	Years after decision			
	1	**2**	**3**	**4**
Panel A.	**Married appellant labor earnings ($1000)**			
Allowed DI	-4.924	-0.917	-4.686	-4.387
	(3.503)	(4.132)	(4.042)	(3.831)
Dependent mean	15.006	14.800	14.201	13.563
Panel B.	**Married appellant total transfers ($1000)**			
Allowed DI	9.478**	6.896	5.392	5.752
	(3.868)	(4.265)	(3.561)	(3.627)
Dependent mean	16.614	17.342	17.905	18.468
Panel C.	**Appellant spouse labor earnings ($1000)**			
Allowed DI	-5.963	-18.305**	-16.166*	-17.806**
	(8.627)	(8.777)	(8.290)	(8.328)
Dependent mean	40.927	39.472	38.751	37.442
Panel D.	**Appellant spouse total transfers ($1000)**			
Allowed DI	0.170	6.241*	6.307	8.620*
	(3.292)	(3.601)	(4.178)	(4.608)
Dependent mean	11.212	11.958	12.654	13.404
Observations	7,813	7,699	7,594	7,480

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

Note: This table displays the impact of DI allowance on earnings and total transfers of married appellants (panels A and B) and their spouses (panels C and D). Baseline estimation sample consists of married appellants who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 different judges. All regressions include fully interacted year and department dummies dummy variables for month of appeal, county of residence, age at appeal, household size, 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.

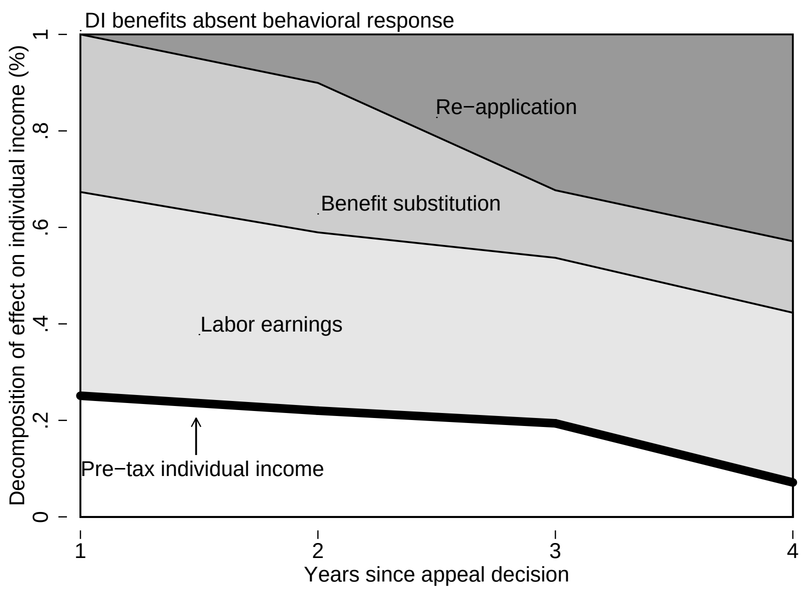
5.4 Attenuation from Reapplication, Benefit Substitution and Labor Earnings

Our estimates above reveal that households respond to DI denials along several margins: increasing household labor supply, substituting towards other benefits programs, and reapplying for DI. We can combine the information in Tables 5-7 to build a comprehensive picture of the extent to which behavioral responses at the individual and household level attenuate the loss in individual benefits stemming from DI denials.

Figure 5 focuses on the attenuation from the reapplication, benefit substitution and labor earnings of the appellant, ignoring for the moment spousal responses. The top horizontal line of the figure corresponds to the effect of DI allowance on individual (pre-tax) income in the first year after the appeal decision (i.e., the gain in income absent any behavioral response, including reapplication, benefit substitution and labor supply). This initial income shock averages approximately $16,000

(Table 6). We normalize it to one in the figure so that all other margins of response are scaled proportionately to this direct effect. The dark grey shaded area below the horizontal line shows the attenuation of the relative income gain of allowed versus denied appellants due to successful reapplications among those initially denied. The medium gray shaded area immediately below corresponds to the attenuation of the income differential stemming from benefit substitution away from other transfer programs. The third shaded area (light gray) plots the behavioral effect of DI allowances on the appellants' labor earnings. Finally, the white area towards the bottom of the figure (below the thick line) shows the net effect of a DI allowance on pre-tax individual income relative to the size of the initial income shock. Accounting for reapplication, benefit substitution, and labor supply, the net effect of a DI allowance on individual income is only 10 to 25 percent as large as the gross DI transfer.

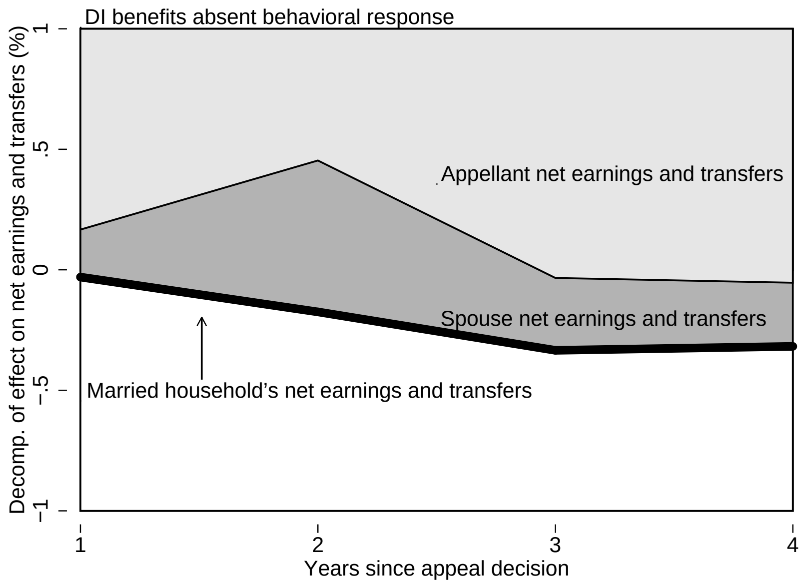
Figure 5: **Decomposing The Effect of DI Allowance on Individual Income**



Notes: This figure uses the IV estimates to decompose the effect of DI allowance on individual income. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). The top horizontal line equals the gain in income from DI allowance in the first year, absent any behavioral response, normalized to one. The triangular area below shows the attenuation of the income gain due to reapplication. The middle area is the attenuation coming from benefit substitution, and the light grey area is the attenuation from individual labor supply response. Initial income shock of DI allowance equals approx. $16,000.

Figure 6 performs an analogous decomposition for the impact of DI allowance on income from labor earnings and net transfers of married couples. To keep this figure legible, we collapse behavioral responses to the DI determination into two categories: appellant responses (labor earnings and transfer payments from reapplication and benefit substitution) and spousal responses (labor earnings and transfer payments from program participation). As per the prior figure, the gross effect of the DI allowance on the income of married couples (averaging $11,000) is normalized to one and plotted using a horizontal line at the top of the figure. The dark gray area immediately below shows that the net effect of appellant reapplications, earnings responses, and benefits substitution. The light gray area below the individual response encompasses the net effect of spousal responses. In combination with appellant level responses, spousal responses more than offset the the loss in DI benefit payments from DI denial. Thus, married appellants who are denied DI do not see a reduction in total household income from labor earnings and transfers.

Figure 6: **Decomposing the Effect of DI Allowance on Total Income from Labor Earnings and Transfers of Married Couples**



Notes: This figure uses the IV estimates to decompose the effect of DI allowance on the net income from labor earnings and transfers of married couples. Baseline estimation sample consists of married individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). The top horizontal line equals the the gain in income from DI allowance in the first year, absent any behavioral response, normalized to one. The dark grey area illustrates the attenuation in income coming from individual response (including earnings, reapplication and benefit substitution), and the light grey area is the spousal response (again, coming from earnings and total transfers). Initial income shock of DI allowance equals approx. $11,000.

6 Consumption Benefits and Fiscal Costs of DI Allowances

In this section, we apply the consumption data detailed in Section 3 to compare the income and consumption gains that DI appellants obtain from DI allowances with the fiscal costs that taxpayers bear, inclusive of DI transfer payments, benefit substitution to or from other transfer programs, and induced changes in tax receipts.

6.1 Pooled estimates

Table 8 reports 2SLS estimates of the impact of allowances versus denials at appeal on per capita household disposable income and consumption. Panel A uses the entire sample of appellants, whereas Panel B uses a restricted sample for whom measurement error in the consumption measures is minimized.[24] To conserve statistical power, we take means of disposable income and consumption at the household level over post-determination years zero through four.

On average, DI allowances raise household consumption (per household member) by approximately \$4,000 per annum. The effect on household income is broadly similar to the consumption impact, suggesting that DI allowance has little impact on household savings. Reassuringly, when we use the restricted sample selected to reduce measurement error, we obtain quite similar estimates for the impact of DI allowances on disposable income and consumption.[25] DI allowances increase both disposable income and household consumption by roughly 15 percent relative to their sample means. Thus, net of the behavioral responses documented above, DI denial imparts meaningful income and consumption gains to individuals at the margin of program entry, as well as their families.

Table 8: **Effects of DI Allowance on Household Disposable Income and Consumption**

	A. Full sample		**B. Restricted sample**	
	Yearly disp. income (per capita)	Yearly consumption (per capita)	Yearly disp. income (per capita)	Yearly consumption (per capita)
Allowed DI	2.786*	3.660*	3.863**	4.266
	(1.468)	(2.152)	(1.900)	(2.716)
Dependent mean	26.195	28.701	25.374	26.219
Observations	14,077	14,077	10,933	10,933

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

Note: This table displays the impact of DI allowance on household disposable income and consumption per household member for the full sample (Panel A) and the restricted sample (Panel B). The dependent variables are measured as average outcomes over the first five post-determination years (including the year of decision). Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 different judges. All regressions include fully interacted year and department dummies dummy variables for month of appeal, county of residence, age at appeal, household size, 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.

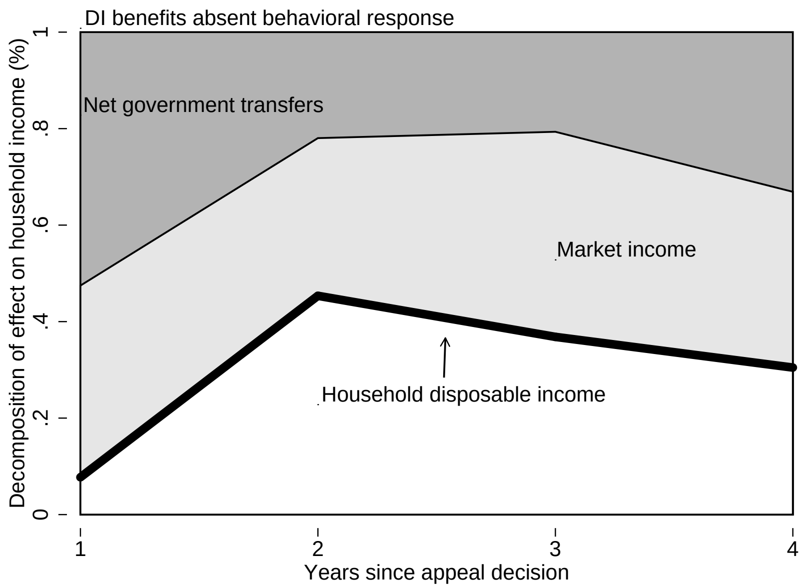
Figure 7 decomposes the effect of DI allowance on disposable household incomes into subcomponents due to net government transfers (total transfers net of taxes) and net market incomes (labor earnings and capital income). As before, we normalize the causal effect of DI allowances on gross DI transfer payments to unity in the base year, so that the effects of transfers and market incomes may be expressed in percentages relative to this value. Approximately 40 percent of the gross

[24] The restricted sample drops observations for which measurement error is more likely, including households with negative income, significant income or assets abroad, and negative net worth in the beginning of a year and zero annual income. This sample also excludes observations in the year in which households buy/sell real estate or divorce/marry, as these events may create errors in the measurement of assets. Brinch *et al.* (2015) provide a detailed discussion of measurement issues in the consumption measure.

[25] As shown in the third column of Appendix Table A.4, there is no evidence of a significant effect of judge leniency on the likelihood of being excluded, which suggest that our estimates based on the restricted sample will not be biased due to endogenous compositional changes.

DI benefit differential between denied versus allowed appellants is offset by large increases in net market incomes of denied relative to allowed appellants. In addition, increases in net government transfer payments to denied versus allowed appellants offset another 20 to 50 percent of the gross differential.

Figure 7: **Decomposing the Effect of DI Allowance on Household Disposable Income**



Notes: This figure uses the IV estimates to decompose the effect of DI allowance on household disposable income. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). The top horizontal line corresponds to the effect of the DI allowance on receipt of DI income in the first year (i.e., the gain in income absent any behavioral response), normalized to one. The dark grey area illustrates the attenuation in household income from other government transfers (including reapplication behavior of denied appellants), and the light grey area is the total attenuation coming from labor market earnings, capital and other sources of income. Initial income shock of DI allowance equals approx. $16,000.

Since we have documented that DI allowances significantly increase disposable income and consumption among appellants and also generally suppress household labor supply, we can infer that DI allowances must have substantial fiscal costs. Table 9 provides a direct accounting of these costs by summing the impact of DI allowances on DI transfer payments, benefit substitution to or from other transfer programs, and induced changes in tax receipts. We estimate that each DI allowance granted on appeal increases net government spending by $16,000 per annum.

To compare these costs to the income gains of the appellants, we multiply the estimated effect per household member in Table 8 by the average number of members per household. The second (and fourth) column of Table 9 reports the ratio of the impact on household income relative to the net effect of fiscal costs. We estimate that DI allowances raise household income by somewhere between 44 and 63 cents per dollar of net government expenditure, with the remainder explained by induced reductions in labor supply that partially offset the increase in transfer income.[26]

[26]Fiscal costs are equal to transfer income minus taxes, while household income is equal to transfer income minus

Table 9: **Effects of DI Allowance on Fiscal Costs**

	A. Full sample		B. Restricted sample	
	Yearly fiscal costs (per allowed)	Benefit-to-cost ratio: ΔHH income/ ΔFiscal cost	Yearly fiscal costs (per allowed)	Benefit-to-cost ratio: ΔHH income/ ΔFiscal cost
Allowed DI	16.475***	0.44	15.631***	0.63
	(4.408)		(4.784)	
Dependent mean	19.611		21.529	
Observations	14,077	14,077	10,933	10,933

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

Note: This table displays the impact of DI allowance on fiscal costs for the full sample (Panel A) and the restricted sample (Panel B). The dependent variables are measured as average outcomes over the first five post-determination years (including the year of decision). Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 different judges. All regressions include fully interacted year and department dummies dummy variables for month of appeal, county of residence, age at appeal, household size, 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. The benefit-to-cost ratio equals the ratio of the effect of DI allowance on total household disposable income over fiscal costs.

6.2 Heterogeneity in Consumption Benefits and Fiscal Costs

Since households differ in their consumption smoothing opportunities, we expect the effects of DI allowances on labor supply, income and consumption to vary across appellants. One element of heterogeneity is household structure. Our evidence so far suggests that married couples may be better equipped than single appellants to offset the financial consequences of denials. In Tables 10 and 11, we split the sample of appellants according to marital status where, as expected, we find stark contrasts. DI allowances generate very large positive impacts on disposable income and consumption among single and unmarried appellants, raising each outcome by approximately 40 percent relative to baseline. These estimates imply that each net dollar in public expenditure induced by a DI allowance raises household income of single and unmarried awardees by nearly 90 cents.[27]

For married couples, accounting for the effect of DI allowances on household labor supply and net payments across all public transfer programs substantially alters our picture of the consumption benefits and fiscal costs of disability receipt. As shown in Panel B, DI denials do not decrease the household income or consumption of married applicants—instead, household labor supply and benefit substitution entirely offset the loss in DI benefit payments. DI allowances to married applicants nevertheless come at considerable fiscal cost to other taxpayers in the form of increased cash

taxes plus labor income and other market income (e.g. capital income). Since capital income plays a small role in the DI appellant sample, most of the offset is due to labor supply impacts.

[27] As above, we multiply impacts on disposable income per household member by household size so that both income gains and fiscal costs are scaled on a per-allowance basis.

transfers and reduced payroll tax revenues. Consequently the ratio of household income gains to fiscal transfer costs is negative: appellants in denied households see a modest increase in income relative to those in allowed households, pointing to discrete choices in labor supply (e.g., due to fixed costs associated with working).

Table 10: **Effects of DI allowance on Household Disposable Income and Consumption**

	A. Unmarried and single		**B. Married**	
	Yearly disp. income (per capita)	Yearly consumption (per capita)	Yearly disp. income (per capita)	Yearly consumption (per capita)
Allowed DI	9.086***	9.835*	-1.615	-0.830
	(3.132)	(5.340)	(2.077)	(2.892)
Dependent mean	24.857	25.934	25.681	26.256
Observations	4,993	4,993	5,929	5,929

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

Note: This table displays the impact of DI allowance on average household disposable income and consumption for the restricted sample of single and unmarried appellants (Panel A) and the restricted sample of married appellants (Panel B). The dependent variables are measured as average outcomes over the first five post-determination years (including the year of decision). Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 different judges. All regressions include fully interacted year and department dummies dummy variables for month of appeal, county of residence, age at appeal, household size, 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.

Table 11: **Effects of DI Allowance on Fiscal Costs**

	A. Unmarried and single		**B. Married**	
	Yearly fiscal costs (per allowed)	Benefit-to-cost ratio: ΔHH income/ ΔFiscal cost	Yearly fiscal costs (per allowed)	Benefit/-to-cost ratio: ΔHH income/ ΔFiscal cost
Allowed DI	19.567***	.86	13.661	-.37
	(6.650)		(8.670)	
Dependent mean	25.566		11.878	
Observations	4,993	4,993	5,929	5,929

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

Note: This table displays the impact of DI allowance on average fiscal costs for the restricted sample of single and unmarried appellants (Panel A) and the restricted sample of married appellants (Panel B). The dependent variables are measured as average outcomes over the first five post-determination years (including the year of decision). Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 different judges. All regressions include fully interacted year and department dummies dummy variables for month of appeal, county of residence, age at appeal, household size, 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. The benefit-to-cost ratio equals the ratio of the effect of DI allowance on total household disposable income over fiscal costs.

While spousal labor supply is a central margin of response to the outcome of DI appeals, appellants may also differ in their own labor supply responses or benefit substitution following the DI determination. To examine heterogeneity in responses according to observable characteristics, we have performed 2SLS estimation of equations (1) and (2) separately for the subsamples in Appendix Table A.3, where we split the sample according to disorder, gender, age, education, and pre-appeal levels of liquid assets. Unfortunately, these subsample estimates are too imprecise to draw any clear conclusions.

To examine evidence for heterogeneity in the effects of DI allowance according to unobserved disability severity (i.e., disability severity not accounted for by the covariates X), we follow French & Song (2013) and Maestas *et al.* (2013) in estimating Marginal Treatment Effects (MTE) of DI receipt. These MTE estimates tell us whether the effects of DI allowance is larger or smaller among appellants who are allowed DI by a strict judge as compared to those who are allowed DI by a lenient judge. Under the monotonicity assumption, the former group of appellants will on average be more severely disabled than the latter group of appellants. The MTE therefore tell us how the effects of DI allowance vary with the degree of (unobserved) appellant severity.

The first step in our computation of the MTE is to use a probit model to estimate the probability of being allowed DI as a function of the instrument and the vector of covariates. The predicted DI allowance rate from this model allows us to define the values of unobserved appellant severity over which the MTE can be identified. Appendix Figure A.9 shows the predicted DI allowance rate in the samples of allowed and denied appellants, where a lower predicted allowance probability corresponds to a stricter judge. An appellant allowed at a lower value of the probability index is therefore expected to have a higher degree of unobserved severity. The figure underscores that our instrument gives little support outside the interval (0.01, 0.30), meaning that we are able to identify the MTE for appellants between the 70th and the 99 percentile of the distribution of unobserved appellant severity.

The second step is to compute numerical derivatives of a smoothed function relating mean household income (or fiscal costs) to the predicted DI allowance rate. In this step, we use the semi-parametric method of Heckman *et al.* (2006; page 8 of web appendix) and restrict the estimation sample to observations for which there is common support. Appendix Figure A.10 shows the estimated MTE as a function of the predicted DI allowance rate, along with bootstrapped 90 percent confidence intervals. The estimated MTE on household income (per household member) and fiscal costs (per allowed) decline monotonically as the predicted allowance rate rises and unobserved severity falls. Concretely, at the righthand side of this index, we see no evidence of a positive impact on a DI allowance on household disposable income, suggesting that denied appellants with less severe disorders offset benefit losses with increased employment and benefit substitution. Conversely, appellants with relatively severe disorders (low values of the index) appear to suffer large net income losses if denied benefits.

7 Deriving Welfare Implications using a Structural Model

We now apply the data and findings above to estimate a simple structural model of household labor supply that translates employment decisions of both spouses into revealed preferences for leisure and consumption. We use this estimated model to compute the welfare benefits of DI receipt—by which we mean the cash equivalent value of receiving a DI allowance at appeal—and to perform counterfactual simulations that allow us to infer the extent to which spousal labor supply and reapplication attenuate the loss in household welfare from DI denial at appeal.

7.1 Description of model and estimation procedure

The cash equivalent value of a DI allowance depends on the preferences for leisure and consumption of appellants and, if relevant, their spouses. To estimate these preferences, we develop both a simple static model and a more realistic dynamic model of household labor supply. For brevity, this section focuses on the static model, while Appendix B describes the dynamic model. At the outset, it is important to emphasize that our goal is to understand the post-appeal labor supply decisions of appellants and their spouses depending on their DI status, and we do not attempt to model the household's behavior prior to the outcome of the appeal decision. Given the outcome of the appeal decision, the appellants and, if relevant, their spouses make labor supply decisions to maximize household utility taken as given their characteristics and economic circumstances.

To evaluate post-appeal outcomes in the static model, we focus on data only from the fourth post-appeal year, by which time many initially denied appellants have reapplied for and have been allowed DI (see Appendix Figure A.11). Appellants and spouses are thus arguably closer to 'steady-state' vis-a-vis their DI status by this date. In the dynamic model, we use data for each year after the appeal decision and explicitly account for re-application.

Preference specification

We consider a unitary model of the household with non-separable preferences between labor supply and consumption and the utility of leisure among spouses.[28] The household utility function depends on disposable income (per household member) I, an employment indicator for the appellant $P_A \in \{0,1\}$, an employment indicator for the spouse $P_S \in \{0,1\}$, and a set of demographic characteristics that may shift preferences K.[29] We abstract from savings, and assume that households consume their disposable income in a given period. This simplifying assumption is consistent with the strong similarity in the estimated effects of DI allowance on household consumption and income, and it fits well the relatively low levels of liquid assets the households hold prior to the appeal.

We follow Blundell *et al.* (2015) in the specification of the preferences. We write the utility function of single and non-married individuals as:

$$U_1(I, P_A; K) = \frac{I^{\mu_1}}{\mu_1} exp\{P_A K'\beta_1\}, \tag{5}$$

where K includes an indicator variable for the appellant being female, an indicator variable for having children under the age of 18, and a constant. The bracketed expression reflects how the marginal utility of income changes with working, according to family demographics; it is normalized to zero if the appellant is not working. To ensure that utility is increasing in consumption, $P_A K'\beta_1 + \mu_1$ is constrained to be negative. To make sure that utility is decreasing in labor supply, $P_A K'\beta_1$ is constrained to be positive.

We assume that the utility function takes a comparable form for married couples, with the addition that preferences depend non-separably on the labor supply of both spouses:

$$U_2(I, P_A, P_S; K) = \frac{I^{\mu_2}}{\mu_2} exp\{P_A K'\beta_A + P_S K'\beta_S + P_A P_S K'\beta_{AS}\}, \tag{6}$$

Analogous to above, we normalize the bracketed expression in equation (6) to zero if both spouses are not working, and we impose the constraints that utility is increasing in consumption and decreasing in labor supply. These constraints require that the bracketed expression is positive and that the

[28] This flexible specification of preferences accommodates non-market production and work related expenses (see e.g. Aguiar & Hurst, 2013). For evidence on non-separability between labor supply and consumption, we refer to Browning & Meghir (1991) and Blundell *et al.* (1994).

[29] As in Maestas *et al.* (2013) and Kostol & Mogstad (2014), employment is an indicator variable that is equal to one if annual earnings exceed the annual substantial gainful activity threshold, set annually by the Norwegian Social Security Administration (at approximately USD 12,500 per year). We are unable to measure labor supply at the intensive margin because we lack reliable data on working hours.

sum of μ_2 and the bracketed expression is negative.

Budget constraint

Appellants and their spouses make labor supply decisions to maximize household utility taken as given their characteristics and budget constraint. For both married and non-married DI appellants, we assume that the income available for consumption depends on the employment decisions of the appellant and the appellant's spouse if relevant, the DI status of the appellant, and a vector of observable characteristics M, which includes the past earnings and prior labor market experience of each spouse. These characteristics are measured prior to the appeal decision, and are supposed to capture heterogeneity in skills and abilities that affect potential earnings.

For each combination of (D,P_A), the potential disposable income of a single or non-married appellant is approximated by the expression:

$$I(D, P_A, M) = M'\delta_{D,A} + v_1,\ \forall(D, P_A) \in \{0, 1\}^2. \tag{7}$$

Similarly, for each combination of (D,P_A,P_S), the potential disposable income of a married couple is approximated by:

$$I(D, P_A, P_S, M) = M'\delta_{D,A,S} + v_2,\ \forall(D, P_A, P_S) \in \{0, 1\}^3. \tag{8}$$

From the parameters $(\delta_{D,A}, \delta_{D,A,S})$, we construct the budget constraints of the households. By linking DI status and employment decisions to disposable income instead of earnings, equations (7) and (8) sidestep the difficulties in specifying the entire tax-benefit system in Norway.

Estimation and identification

By comparison to the IV model, identification of the structural model relies on strong functional form and exogeneity assumptions. For example, gender and parental status shift preferences but do not affect potential income, whereas past earnings and prior labor market experience affect potential income but do not shift preferences. As a result, we get variation in potential income for people with the same preferences, helping to identify the model. On top of this, the randomness in the disability determination process gives exogenous variation in DI status and, therefore, also in the income available for consumption.

We use the method of simulated moments to estimate the parameters of the model.[30] We estimate the model separately for married couples (9 preference parameters and 40 budget constraint parameters) and single and non-married individuals (3 preference parameters and 12 budget

[30] All parameters are estimated except for μ_1 and μ_2. Following Blundell *et al.* (2015), we set these parameters to be equal to -0.56. In the static model, this is a free normalization, implying that neither the welfare values, the model fit, or the labor supply elasticities are affected. In the dynamic model, the choice of μ_1 and μ_2 matter for the welfare calculations. The chosen value of these parameters correspond to a risk aversion coefficient of 1.56, consistent with the evidence cited in Blundell *et al.* (2015).

constraint parameters).

In the sample of single and non-married individuals, the parameters $(\delta_{D,A}, \beta_1)$ are jointly estimated by matching (a) the IV estimates of DI participation on D and P_A with judge leniency as the instrument, and (b) the covariance matrix between I and Q for each combination of (D,P_A), where Q includes K, M, and dummy variables for education levels of the appellant and the number of children.[31] Similarly, for the sample of married couples, we jointly estimate the parameters $(\delta_{D,A,S}, \beta_A, \beta_S, \beta_{AS})$ by matching (a) the IV estimates of DI participation on D,P_A, and P_S with judge leniency as the instrument, and (b) the covariance matrix between I and Q for each combination of (D,P_A,P_S), where Q includes K, M, and dummy variables for education levels of both spouses and the number of children.

Our estimation procedure finds the parameters that minimize the differences between the simulated moments and the actual moments.[32] In both samples, we assign equal weight to (a) and (b). Within (a) and (b), our estimation procedure minimizes the weighted sum of squared residuals, with weights given by the inverse of the standard deviations of the estimated moments. Taken together, there are 107 moments to match for married couples and 38 moments to match for single and non-married individuals.

Brief description of the dynamic model

A key limitation of our static model is that it abstracts from reapplication, which is prevalent and consequential in the data. Hence, we also present results below from a dynamic version of the model, detailed in Appendix B. The dynamic model uses data from each post-appeal year and includes a stochastic reapplication process to reproduce the patterns observed in the data. In particular, the year-to-year transitions to DI are approximated by a probit model where the probability of receiving DI in year t depends on DI participation in year $t-1$, employment decisions in year $t-1$, and M. In each period, the household maximizes expected lifetime utility taken as given its current characteristics, the budget constraint, and the DI transition probabilities. Individuals therefore take account of how working today may lower their chances of being allowed DI in the future (e.g. because it signals work capacity).

Parameters of both the static and dynamic models are estimated by matching the IV estimates of DI participation on employment and household income as well as the full covariance matrix between disposable income and family demographics for every combination of DI status and employment decision, as described above. The key difference is that the static model targets the moments in year four only, whereas the dynamic model matches the moments in each of the four years after the appeal decision.

[31] The fact that Q contains additional heterogeneity that is not explicitly modeled in the specification of preferences or the budget set is diagnostic: these additional degrees of freedom allow the model to poorly match the moments if it badly approximates the underlying data generating process.

[32] We use the BFGS (Broyden–Fletcher–Goldfarb–Shanno) numerical optimization algorithm. Our parameter estimates appear robust to the choice of starting values.

7.2 Empirical results

Model fit and parameter estimates

Table 12 compares the IV estimates of the causal effects of DI participation on employment and disposable income in year four with the model simulated effects of DI participation on these variables. These IV estimates are obtained, as above, from 2SLS estimation of equations (1) and (2), except that now, the endogenous regressor is an indicator for whether an appellant is participating on DI in a given year (rather than whether the appellant was allowed DI at the appeal). This specification of the IV model conforms to the structural model, where employment decisions and disposable income depend on contemporaneous DI status.[33] For both unmarried individuals and married couples, the structural estimates from both the static and dynamic models closely replicate the IV estimates for the impact of DI participation on employment and household income in year four.

Table 12: **Fit of Instrumental Variables Estimates, Year 4**

	Employment		**Household income**
	Appellant	**Spouse**	**($1000, per capita)**
Panel A.		**Married**	
IV estimate of effect of DI participation	-0.191	-0.663	-2.064
Simulated effect of DI participation			
Static model	-0.166	-0.641	-1.774
Dynamic model	-0.183	-0.584	-2.172
Panel B.		**Unmarried and single**	
IV estimate of effect of DI participation	-0.269		9.131
Simulated effect of DI participation			
Static model	-0.265		8.872
Dynamic model	-0.261		8.834

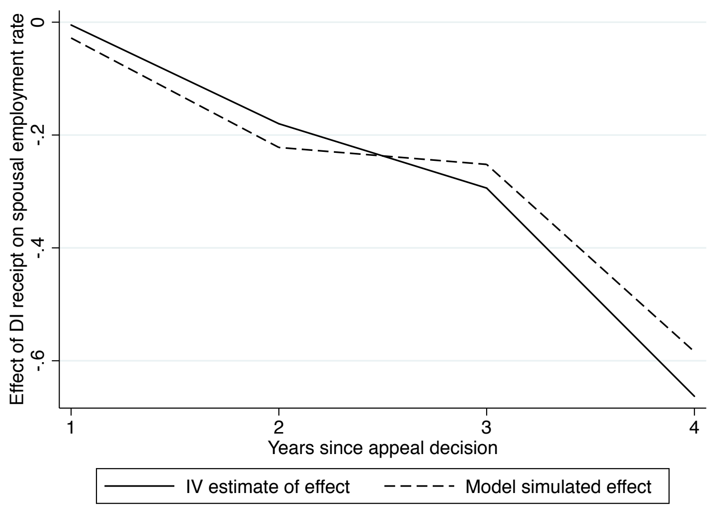
Notes: This table compares the IV estimates of DI participation on employment and disposable income in year four to the model simulated effects of these variables.

In estimating the dynamic model, we match the IV estimates on employment and disposable income in each of the four years after the appeal decision. Figure 8 compares the IV estimates of how DI participation affects spousal employment over time to the model simulated effects on these variables. Importantly, the dynamic model is able to replicate the time trend in the effects of DI participation on spousal employment. When we consider the other outcomes (which vary much less over time), the dynamic model also performs well. Appendix Figure A.12 summarizes

[33] The reduced form of the IV model is, of course, unaffected by whether we use initial allowance or DI participation as the endogenous regressor. What differs is the first stage estimate, which is lower for DI participation than for initial allowance because of reapplication among initially denied appellants.

how the models fit all moments (including the IV estimates and the covariance matrix). Since the variables behind the moments are measured in different units (e.g. income versus employment), we divide the difference between an actual and a simulated moment by the standard deviation of the respective variable. Appendix Figure A.12 displays the distribution of these standardized differences. The graphs show that the model fits the data rather well, both for married couples and unmarried individuals. The distributions are centered around zero, and we rarely observe differences that exceed 0.2 standard deviation of the variables.

Figure 8: **Dynamic Model: Fit of Spousal Employment Over Time**



Notes: This figure compares the IV estimates of the effects of DI participation on spousal employment rates to the model simulated effects on these variables.

Appendix Table A.5 presents the estimates of the preference parameters. The estimated intercept terms are always positive, implying that employment reduces the value of consumption. The disutility from work tends to be more pronounced for women then for men, and is considerably more negative for unmarried individuals with children than unmarried individuals without children. The coefficients on the interactions between the employment of the spouses are positive, implying that the presence of a working spouse further increases the cost of working. This suggests some degree of substitutability between the labor supply of spouses.

Using the fitted models to simulate how employment rates of appellants and spouses change with a one percent increase in disposable income from working, we obtain plausible labor supply responses. Because few appellants on DI are working, we focus on employment responses in the non-DI state. In year four, for example, the preference parameters from the dynamic model imply a 0.99 percentage point change in the employment rate of spouses per one percent increase in disposable income from working. This translates into an own-income employment elasticity of 1.59. By comparison, the employment rate of married appellants is predicted to increase by 0.36 percentage points for a one percent increase in household disposable income. This corresponds to an own-income employment elasticity of 1.43. These sizable employment responses are within the range of those reported in Blundell *et al.* (2015). As expected, the cross-income elasticities are always negative and smaller (in absolute value) than the own-income elasticities. For example, the employment elasticity of the spouse with respect to a one percent change in the appellant's disposable income from working is -0.22 in year four of the dynamic model.

Household valuation of DI receipt

Building on the good fit of the structural estimates to the IV estimates of the effects of DI participation on employment and total household income—and the plausible labor supply elasticities implied by the models—we now apply the structural model estimates to explore households' valuation of DI receipt, by which we mean the amount of disposable income that denied appellants must be compensated to maintain the level of utility when allowed.

We begin with estimates from the static model. As shown in Table 13, both married couples and unmarried individuals have considerable willingness to pay for DI receipt. For married couples, the average willingness to pay is nearly $2,700 per household member. Since the effect of DI participation on disposable income for married couples is weakly negative (see Table 10), the entire utility value of DI receipt can be attributed to increased leisure of appellants and their spouses. We estimate that single and unmarried appellants would be willing to pay considerably more than married appellants for DI receipt. For this group, the average willingness to pay for DI receipt is nearly $9,100 per household member, mirroring closely the estimated effect of DI participation on disposable income. These results imply that DI receipt is more valuable for single and unmarried appellants as compared to married couples. These findings provide a preference-based interpretation of our IV results: DI benefits primarily insure single and unmarried appellants against sharp falls in consumption; they largely insure married appellants and their spouses against losses of individual and joint leisure rather than declines in consumption. Willingness to pay for these goods differs accordingly.

To address the shortcoming that the first three post-appeal years of outcome data—and the corresponding reapplication process—are omitted from the static model, we also apply the dynamic model to compute households' willingness to pay for being allowed DI at the initial appeal. We define willingness to pay as the amount of disposable income that, in every year after the appeal decision, must be added to I when denied DI at the appeal to leave the household at the same level of expected utility as when allowed DI. As shown in Table A.5, the welfare benefits for married couples are broadly similar whether we use the static or the dynamic model. One difference between the two models is that the dynamic model yields a lower implied willingness to pay for DI benefits among single and unmarried appellants than does the static model ($5,900 in the dynamic model versus $9,100 in the static model). Nevertheless, the willingness to pay for DI receipt remains larger for single and unmarried individuals than for married couples, suggesting that it might be efficient to offer lower replacement rates or impose stricter screening on married appellants. Of course, any policy that conditioned disability screening and benefits on marital status would have to account for likely policy-induced shifts in marriage formation and dissolution.

Table 13: **Estimated Household Valuation of DI Receipt**

	Static Model Year 4	**Dynamic Model** Annuity value
Panel A.	**Married**	
Average willingness to pay ($1000, per capita):		
Unconstrained	2.677	2.896
Constraining spousal employment	7.057	4.827
No reapplication		4.478
Constraining spousal employment & no reapplication		7.361
Panel B.	**Single and unmarried**	
Average willingness to pay ($1000, per capita):		
Unconstrained	9.076	5.896
No reapplication		9.108

Note: This table shows estimates of the average welfare benefit ($1000, per household member) of DI receipt for married people and single and unmarried individuals. The first column refer to the static model, while the second column refers to the dynamic model. In the rows titled "Unconstrained", we use the estimated model to compute the welfare benefit of DI receipt. In the row titled, "Constrained spousal employment," we compute the willingness to pay for DI receipt while constraining the spousal labor supply responses to DI receipt by setting the spouse's employment decision when D=0 to be equal to her employment decision when D=1. In the rows titled "No reapplication,"we compute the willingness to pay for DI receipt while constraining denied appellants from reapplying for benefits by setting the probability of transitioning into DI equal to zero.

Quantifying the importance of spousal labor supply and re-application

Lastly, we perform counterfactual analyses that help us infer the extent to which spousal labor supply and re-application attenuate the welfare loss from being denied DI at the appeal. In the first exercise, we prevent any spousal labor supply response to DI receipt by setting the spouse's employment decision when $D = 0$ to be equal to her employment decision when $D = 1$. This means that the spouse is prevented from adjusting her labor supply in response to the appellant being denied DI. Notably, eliminating the option for a spousal labor supply response substantially increases the willingness of married couples to pay for DI. Indeed, this restriction eliminates much of the differences in the willingness to pay of married couples versus single and unmarried individuals. This underscores the importance of spousal labor supply in allowing households to buffer the welfare loss stemming from DI denial.

In a final counterfactual exercise, we impose the constraint that denied appellants cannot reapply for DI benefits. Formally, we set the probability of transitioning into DI equal to zero, so appellants who are denied at the appeal will never receive DI. By comparing appellants' willingness to pay in the constrained and unconstrained settings, we estimate that reapplication attenuates around 35 percent of the the welfare loss from DI denial (relative to a setting where reapplication is infeasible). Taken together, the options for appellants to reapply for disability benefits and for their spouses to increase labor supply jointly reduce the welfare loss associated with DI denial on the first appeal by as much as 60 percent.

8 Conclusion

While a mature literature finds that DI receipt discourages work, the welfare implications of these findings depend on two rarely studied economic quantities: the value that individuals and families place on receipt of disability benefits; and the full cost of DI allowances to taxpayers, summing over DI transfer payments, benefit substitution to or from other transfer programs, and induced changes in tax receipts. We assessed these missing margins in the context of Norway's DI system, drawing on two strengths of the Norwegian environment. First, Norwegian register data allowed us to characterize the household impacts and fiscal costs of disability receipt by linking employment, taxation, benefits receipt, and assets at the person and household level. Second, random assignment of DI applicants to Norwegian judges who differ systematically in their leniency allowed us to recover the causal effects of DI allowance on individuals at the margin of program entry.

Accounting for the total effect of DI allowances on both household labor supply and net payments across all public transfer programs substantially alters our picture of the consumption benefits and fiscal costs of disability receipt. While DI denial causes a significant drop in household income and consumption on average, it has little impact on income or consumption of married appellants; spousal earnings and benefit substitution entirely offset the loss in DI benefit payments. To explore welfare implications, we estimated a structural model of household labor supply that translates employment decisions of both spouses into revealed preferences for leisure and consumption. We used the estimated model to compute the welfare benefits of DI receipt, and to perform counterfactual exercises that help us infer the extent to which spousal labor supply and reapplication attenuate the welfare loss from DI denial at appeal. We found that the welfare benefit of DI receipt is considerably larger for single and unmarried individuals as compared to married couples, suggesting that it might be efficient to lower replacement rates or impose stricter screening on married appellants.

When considering the interpretation and generality of these findings, we emphasize two caveats. First, our structural model permits us to estimate the economic value of the transfer component of DI benefits—that is, the cash equivalent value of a DI award—but does not encompass the ex ante insurance value of the DI system for potential applicants. Since this insurance value is doubtless positive and potentially large, our estimates should not be interpreted as a full accounting of the welfare value of the DI system. Second, the estimates obtained by the quasi-experimental variation in judicial disability determinations correspond to the local average treatment effect of DI allowance or denial for individuals who could potentially have received a different allowance decision in the appeal process had their case been assigned to a different judge. Since the work capacity of individuals at the margin of program entry is likely to differ from that of inframarginal individuals, we are cautious in extrapolating the causal estimates obtained here to the broader population at large or to other programmatic settings.

Nevertheless, the economic consequences of DI receipt for marginal DI claimants are relevant for policy. In both Norway and the U.S., the rise in DI rolls in recent decades is driven in significant part by de jure or de facto changes in the screening criteria applied to claimants reporting difficult-to-verify disorders, such as back pain or mental disorders (Autor & Duggan, 2006; Kostol & Mogstad, 2014). Logically, reforms aimed at altering DI screening criteria are likely to have the largest impacts on applicants on the margin of program entry, a substantial share of whom are applicants with difficult-to-verify disorders. Not coincidentally, this description also corresponds closely to the marginal appellants whose outcomes identify the causal effects estimates and model-based welfare calculations above. These observations suggest that while the estimates provided by this paper are not directly generalizable to the full DI population, they are likely to be informative for policymaking.

References

Aguiar, Mark, & Hurst, Erik. 2013. Deconstructing Life Cycle Expenditure. *Journal of Political Economy*, **121**(3), 437 – 492.

Atkinson, Anthony B., Rainwater, Lee, & Smeeding, Timothy M. 1995. Income Distributions in OECD Countries: Evidence from the Luxembourg Income Study. *OECD Publications and Information Center*.

Autor, D. G., Maestas, N., Mullen, K., & Strand, A. 2014a. Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability. *NBER Working Paper*.

Autor, D. H. 2011. The Unsustainable Rise of the Disability Rolls in the United States: Causes, Consequences, and Policy Options. *NBER Working Paper*.

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., Duggan, M. G., & Gruber, J. 2014b. Moral Hazard and Claims Deterrence in Private Disability Insurance. *American Economic Journal: Applied Economics*, **6**(4), 110–141.

Baicker, Katherine, Finkelstein, Amy, Song, Jae, & Taubman, Sarah. 2014. The Impact of Dedicaid on Labor Market Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment. *American Economic Review*, **104**(5), 322–328.

Blundell, R., Pistaferri, L., & Preston, I. 2008. Consumption Inequality and Partial Insurance. *The American Economic Review*, **98**(5), pp. 1887–1921.

Blundell, R., Pistaferri, L., & Saporta-Eksten, I. 2012. Consumption Inequality and Family Labor Supply. *NBER Working Paper*, October.

Blundell, R., Costa Dias, M., Meghir, C., & Shaw, J. 2015. Female Labor Supply, Human Capital and Welfare Reform. *Discussion Paper*.

Blundell, Richard, Browning, Martin, & Meghir, Costas. 1994. Consumer Demand and the Life-Cycle Allocation of Household Expenditures. *The Review of Economic Studies*, **61**(1), pp. 57–80.

Borghans, L, Gielen, A., & Luttmer, E. 2014. Social support substitution and the earnings rebound: Evidence from a regression discontinuity in disability insurance reform. *American Economic Journal-Economic Policy*, **6**(4), 34–70.

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

Brinch, C., Eika, L., & Mogstad, M. 2015. What can we learn about household consumption from information on income and wealth? *Mimeo*.

Browning, M., & Leth-Petersen, S. 2003. Imputing consumption from income and wealth information. *The Economic Journal*, **113**(488), F282–F301.

Browning, Martin, & Meghir, Costas. 1991. The Effects of Male and Female Labor Supply on Commodity Demands. *Econometrica*, **59**(4), pp. 925–951.

Chen, S., & van der Klaauw, W. 2008. The work disincentive effects of the disability insurance program in the 1990s. *Journal of Econometrics*, **142**(2), 757–784.

Cullen, J., & Gruber, J. 2000. Does Unemployment Insurance Crowd out Spousal Labor Supply? *Journal of Labor Economics*, **18**(3), pp. 546–572.

Dahl, G., Kostøl, A., & Mogstad, M. 2014. Family Welfare Cultures. *The Quarterly Journal of Economics*, **129**(4), 1711–1752.

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

Fadlon, I., & Nielsen, T. H. 2015. Household Responses To Severe Health Shocks And The Design of Social Insurance. *Mimeo*.

Finkelstein, A., Hendren, N., & Luttmer, E. 2015. The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment. *NBER Working Paper No. 21308*, **June**.

French, E., & Song, J. 2013. The Effect of Disability Insurance Receipt on Labor Supply, Appeals and Re-Applications: Evidence from Hearing Office Level Decisions. *American Economic Journal: Applied*, **6**(2), 291–337.

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

Heckman, J. J., Urzua, S., & Vytlacil, E. 2006. Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*, **88**(3), 389–432.

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

Kostol, Andreas, & Mogstad, Magne. 2015. Earnings, Disposable Income, and Consumption of Allowed and Rejected Disability Insurance Applicants. *American Economic Review*, **105**(5), 137–41.

Liebman, Jeffrey B. 2015. Understanding the Increase in Disability Insurance Benefit Receipt in the United States. *Journal of Economic Perspectives*, **29**(2).

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

Lundberg, S. 1985. The added worker effect. *Journal of Labor Economics*, 11–37.

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

Meghir, C., & Pistaferri, L. 2011. Earnings, Consumption and Life Cycle Choices. Handbook of Labor Economics, vol. 4, Part B. Elsevier.

Meyer, B. D., & Mok, W. 2013. Disability, Earnings, Income and Consumption. *NBER Working Paper*, March.

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*.

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

of Trustees of the Federal OASDI Trust Funds, Board. 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.

of Trustees of the Federal OASDI Trust Funds, Board. 2014. *The 2014 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds.* Tech. rept. U.S. Social Security Administration.

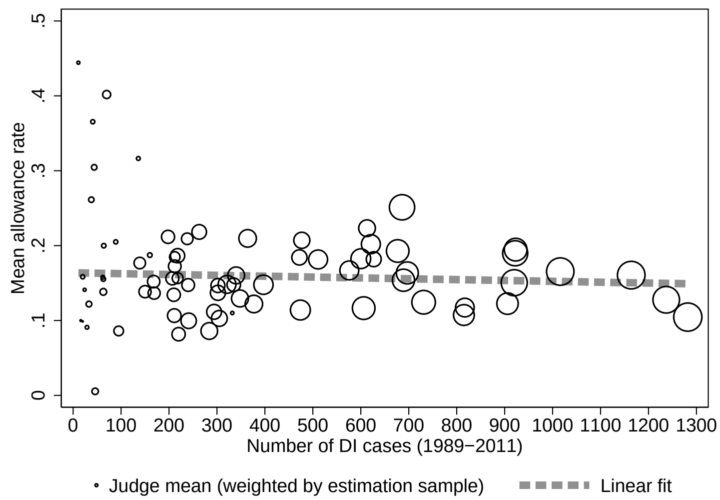
Parsons, Donald O. 1980. The Decline in Male Labor Force Participation. *The Journal of Political Economy*, **88**(1), 117–134.

Pistaferri, L. 2015. Household consumption: Research questions, measurement issues, and data collection. *Mimeo.*

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.

Appendix A. Additional Tables and Figures

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

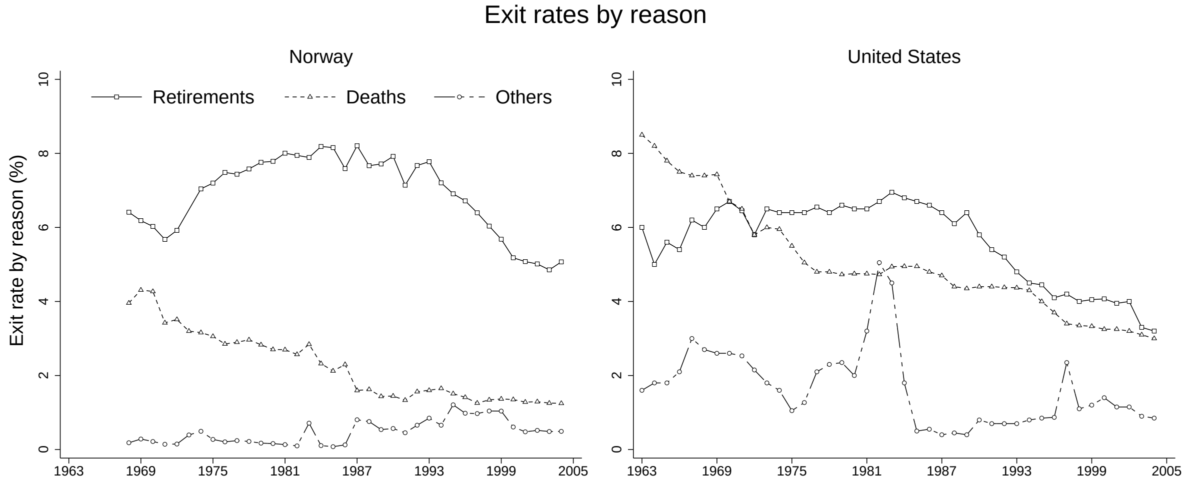


Notes: The figure plots a judge's allowance rate against the total number of cases he or she has handled. There are 75 different judges, and on average, each judge has handled a total of 325 cases. Allowance rates normalized by subtracting off year × department deviations from the overall mean. The sample is restricted to individuals appealing their first denied case during the period 1994-2005. 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: **DI Awards and DI Exits in Norway and the U.S.**

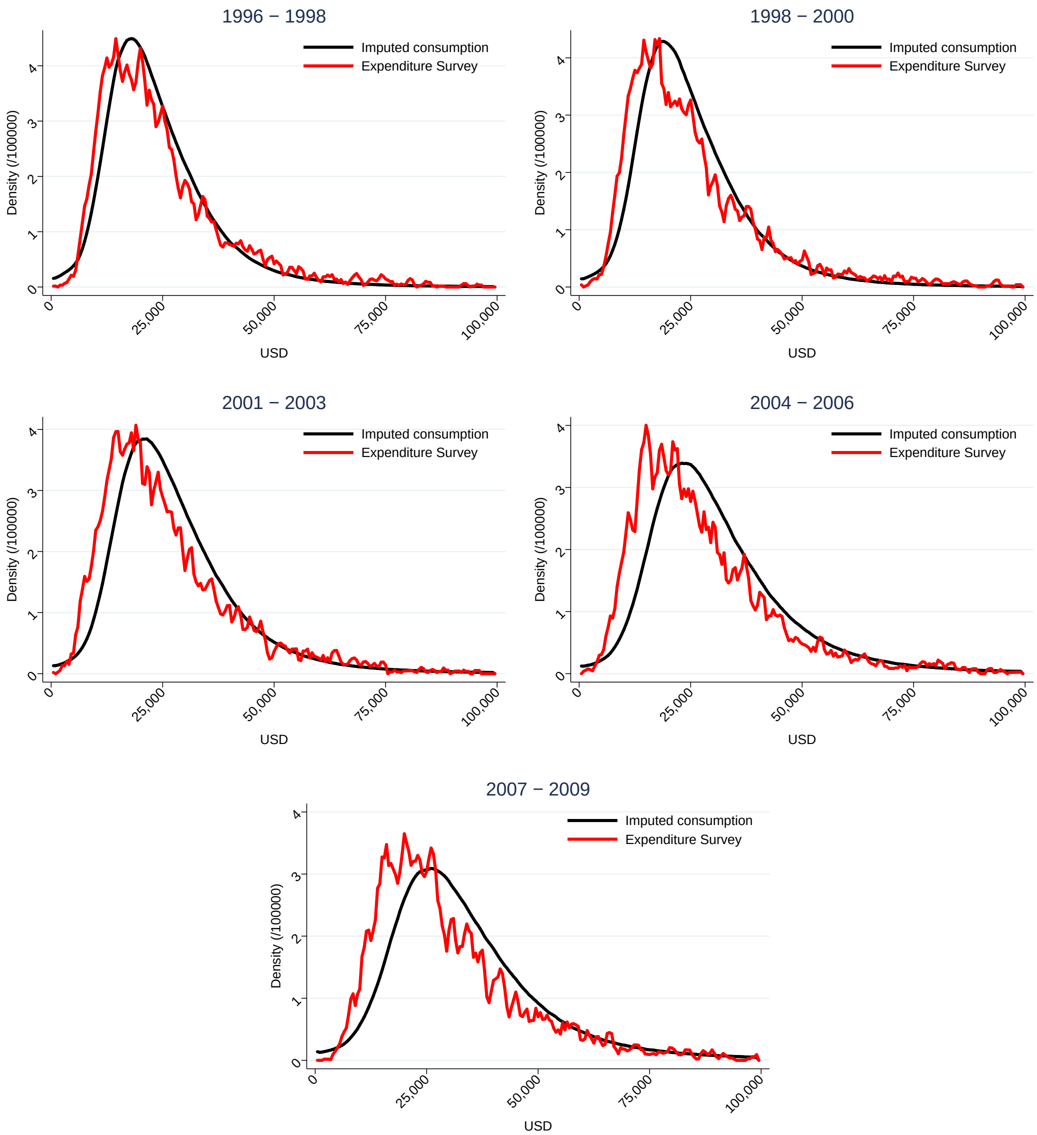
Notes: The U.S. trends are based on Autor & Duggan (2006), while the Norwegian trends are collected from various issues of the SSA Supplement. The graphs show award rates in the insured population and exit rates from the DI program in both countries.

Figure A.3: **DI Exits by Reason in Norway and the United States**



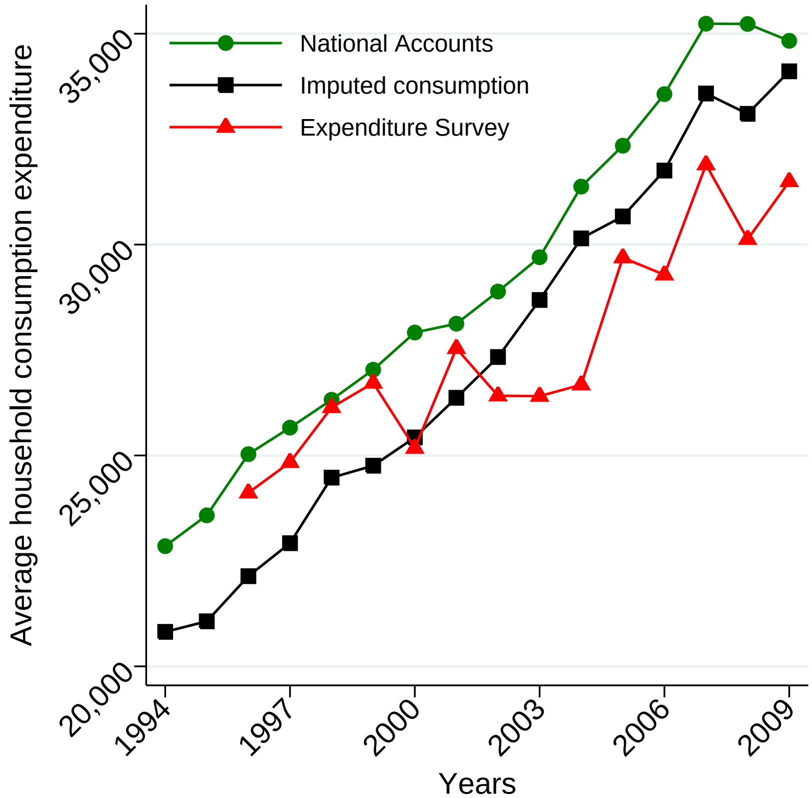
Notes: The U.S. trends are based on Autor & Duggan (2006), while the Norwegian trends are collected from various issues of the SSA Supplement. The graphs show exit rates because of death, retirement or other reasons (including eligibility-based exits).

Figure A.4: **Comparison of Consumption Measures Based on Expenditure Surveys vs. Data on Income and Assets**



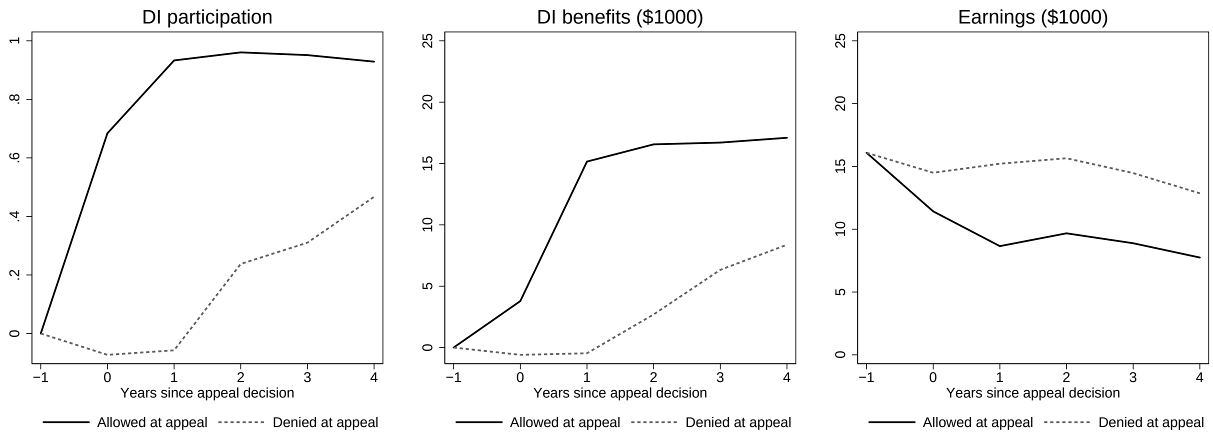
Notes: This figure compares the measures of total expenditure from Norwegian family expenditure surveys to the measures of consumption derived from data on income and assets. Because of measurement issues with the expenditure surveys for 1994-1996, we do not report figures for these years. Source: Brinch *et al.* (2015).

Figure A.5: **Comparison of Aggregate Consumption Expenditure Based on Expenditure Surveys vs. Data on Income and Assets vs. National Accounts**



Notes: This figure compares measures of average household consumption expenditure from Norwegian family expenditure surveys to the measures from national accounts and those derived from data on income and assets. Because of measurement issues with the expenditure surveys for 1994-1996, we do not report figures for these years. Source: Brinch *et al.* (2015).

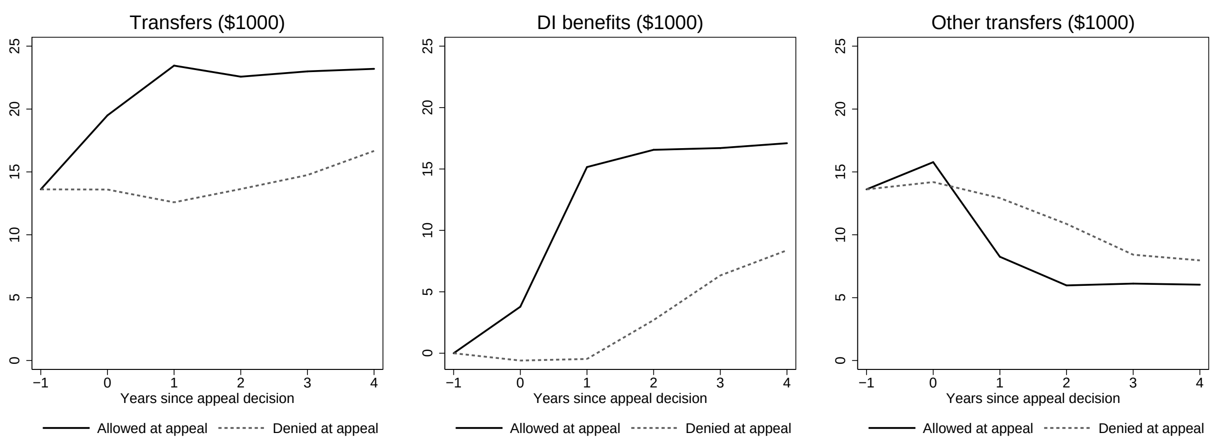
Figure A.6: **Potential Outcomes: Labor Earnings and DI benefits**



(a) (b) (c)

Notes: These figures display the decomposition of our LATE estimates into potential outcomes for allowed and denied complier appellants (see Dahl *et al.* (2014) for details).

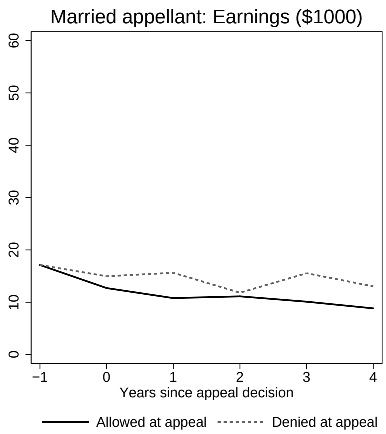
Figure A.7: **Potential Outcomes: Benefit Substitution**



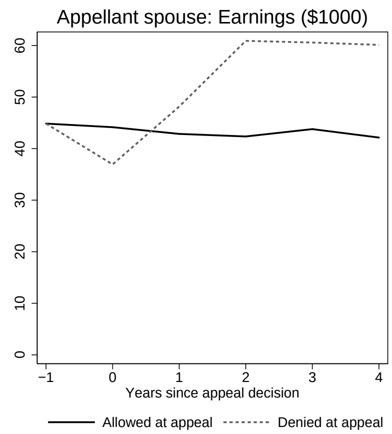
(a) (b) (c)

Notes: These figures display the decomposition of our LATE estimates into potential outcomes for allowed and denied complier appellants (see Dahl *et al.* (2014) for details).

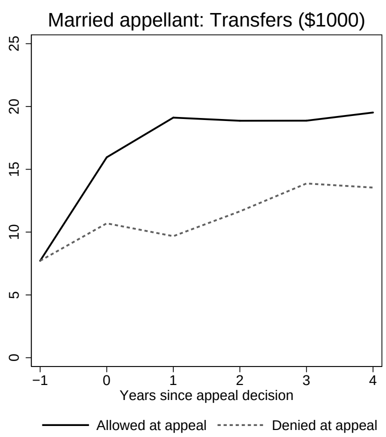
Figure A.8: **Potential Outcomes: Married appellants and Spouses**



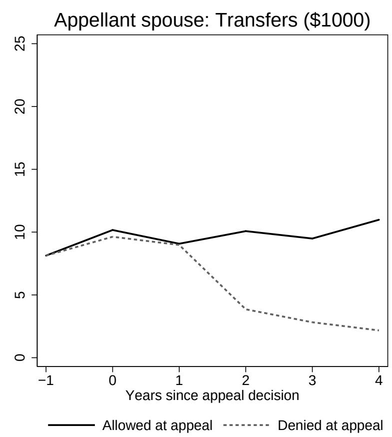
(a)



(b)



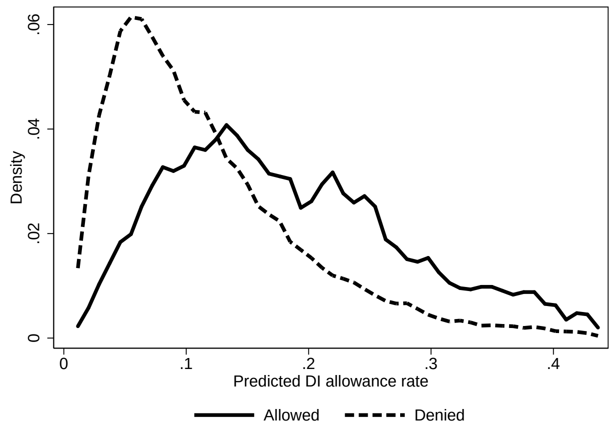
(c)



(d)

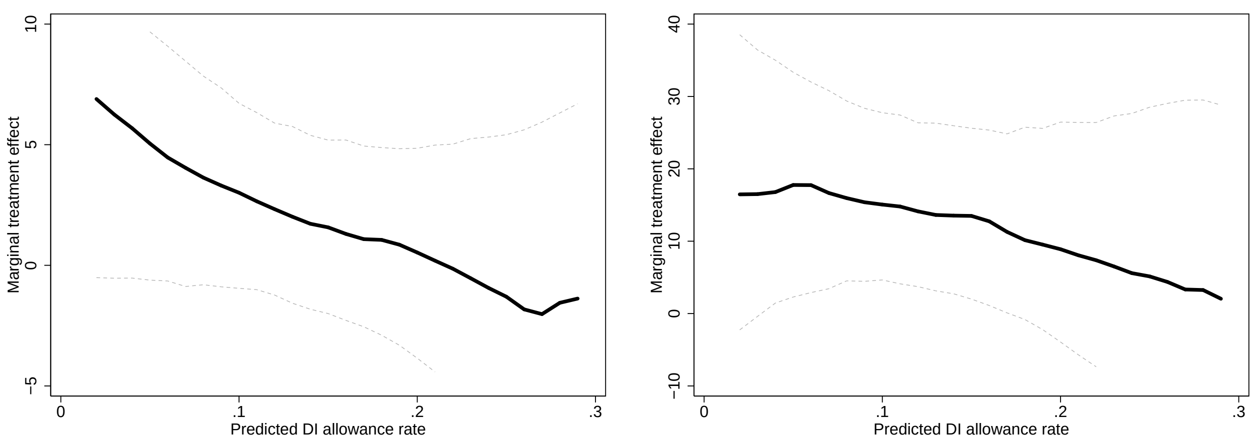
Notes: These figures display the decomposition of our LATE estimates into potential outcomes for allowed and denied complier appellants (see Dahl *et al.* (2014) for details).

Figure A.9: **Predicted Allowance Rate for Allowed and Denied Appellants**



Notes: This figure displays the predicted DI allowance rate in the samples of allowed and denied appellants. The predicted values come from a probit model of DI allowance on the instrument and the covariates.

Figure A.10: **Marginal Treatment Effects on Household Income and Fiscal Costs**

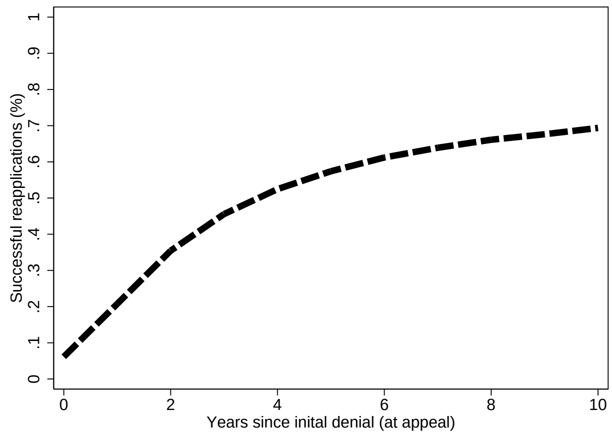


(a) Household disposable income (per capita)

(b) Fiscal costs (per allowed)

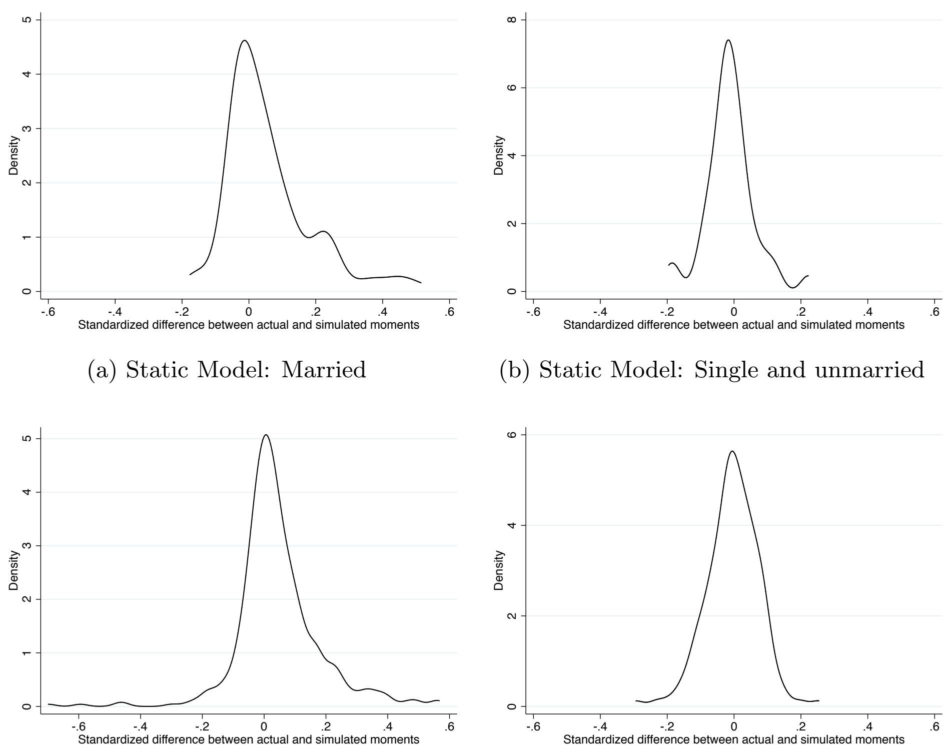
Notes: This figure displays the estimated marginal treatment effects of DI allowance on household disposable income (Panel A) and fiscal costs (Panel B). The sample is restricted to the range of propensity scores with common support, after excluding the top and bottom 1 percentiles. There are 13,741 observations and 75 judges.

Figure A.11: **DI Participation Rate of Denied Appellants Over Time**



Notes: This figure displays the DI participation rate among denied appellants by year since appeal decision.

Figure A.12: **Model Fit: All Moments**



(a) Static Model: Married

(b) Static Model: Single and unmarried

(c) Dynamic Model: Married couples

(d) Dynamic Model: Single and unmarried

Notes: This figure summarizes how well the models fit all the moments (including the IV estimates and the covariance matrix). Since the variables behind the moments are measured in different units (e.g. income versus employment), we divide the difference between an actual and a simulated moment by the standard deviation of the respective variable. The graphs show the distributions of these standardized differences.

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 disorders include 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.

Table A.3: **Sub-Sample First Stage Estimates**

	A. Marriage status		**B. Gender**		**C. Age group**	
	Married	Unmarried	Female	Male	Young	Old
First stage	0.856***	0.737***	0.834***	0.769***	0.773***	0.822***
	(0.094)	(0.099)	(0.084)	(0.116)	(0.079)	(0.107)
Dependent mean	0.135	0.119	0.135	0.116	0.094	0.166
Observations	7,938	6,139	8,844	5,233	7,452	6,625
	D. Diagnosis group		**E. Education level**		**F. Liquid assets**	
	Difficult to verify	Other diagnosis	Low	High	High	Low
First stage	0.810***	0.805***	0.896***	0.714***	0.853***	0.773***
	(0.084)	(0.103)	(0.089)	(0.092)	(0.088)	(0.107)
Dependent mean	0.124	0.136	0.116	0.139	0.144	0.112
Observations	9,806	4,271	7,090	6,987	6,936	7,141

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

Notes: This table displays heterogeneity in first stage estimates. The first stage specification corresponds to Panel B in Table 4. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 different judges. Young (old) age group is defined by below (above) median age 48. High (low) education level is at least (less than) high school degree. High (low) liquid asset is defined based on the medium level of liquid assets ($1600) the year before appeal.

Table A.4: **Specification Checks**

Dependent variable	**Died or migrated**	**Married**	**In restricted sample**
Judge leniency	0.016	0.003	-0.015
	(0.047)	(0.031)	(0.019)
Dependent mean	0.091	0.598	0.981
Observations	14,721	14,077	14,077

***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 on the probability that a person ever die or migrate (first column); is ever married (second column); is ever in the restricted sample (third column). Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). The second and third columns exclude those who die or migrate during the year of the appeal. There are 75 different judges. All regressions mirror the reduced form specification of Table 5.

Table A.5: **Estimated Preference Parameters**

	Static Model	**Dynamic Model**
Panel A.	**Married**	
Appellant working		
Intercept	0.019	0.023
Interacted with female appellant	0.024	0.122
Interacted with child	-0.003	-0.008
Spouse working		
Intercept	0.045	0.061
Interacted with female appellant	-0.021	0.022
Interacted with child	-0.016	0.002
Both spouses working		
Intercept	0.038	0.015
Interacted with female appellant	0.032	0.021
Interacted with child	0.072	-0.052
Panel B.	**Single and unmarried**	
Appellant working		
Intercept	0.005	0.007
Interacted with female appellant	0.000	0.030
Interacted with child	0.525	0.519

Note: This table displays estimates of the preference parameters. The parameters reflect how the marginal utility of consumption changes with working, by family demographics. In the static (dynamic) model, the utility is normalized to zero if the appellant (and the spouse) is not working. The first column refers to the static model, and the second column refers to the dynamic model.

Appendix B: Dynamic Model of Household Labor Supply

In each year after the appeal decision, a unitary household maximizes expected life-time utility taken as given its current characteristics and economic circumstances. We assume that utility is inter-temporally separable. Focusing on the sample of married couples at any year t, the household's problem can be written as

$$\begin{aligned} V_t(D_t; K, M) = \max_{P_{A,t}, P_{S,t}} E\Big\{ & U(P_{A,t}, P_{S,t}, I_t(P_{A,t}, P_{S,t}, D_t, M); K) \\ & + \delta \pi_t(P_{A,t}, P_{S,t}, D_t, M) V_{t+1}(D_{t+1} = 1, K, M) \\ & + \delta(1 - \pi_t(P_{A,t}, P_{S,t}, D_t, M)) V_{t+1}(D_{t+1} = 0, K, M) \Big\}, \; t = 1, 2, \ldots, (T-1), \end{aligned}$$

where E is the expectation operator conditional on the available information at age t over all future random events, δ is the discount rate set equal to 0.98 (as in Blundell *et al.* , 2015), π_t is the transition probability between states (described below), and V_t is the optimized present discounted value of future utility. The instantaneous utility of married couples is given by equation (6). The continuation value is calibrated as,

$$
\begin{aligned}
V_T(D_T;K,M) &\equiv \max_{P_{A,T},P_{S,T}} \sum_{t=\mathrm{Age}_T}^{67-1} \delta^{t-\mathrm{Age}_T} U(P_{A,T},P_{S,T},I_T(P_{A,T},P_{S,T},D_T,M);K) \\
&= \frac{1-\delta^{67-\mathrm{Age}_T}}{1-\delta} U(P_{A,T},P_{S,T},I_T(P_{A,T},P_{S,T},D_T,M);K)
\end{aligned}
$$

where Age_T denotes age at time T, and age 67 is the typical retirement age in Norway. This specification of the continuation value means that we abstract from pension-related work incentives.

Maximization must respect two constraints, the first of which is the budget constraint. We abstract from household formation and dissolution, and we assume no savings, implying that households consume their disposable income in each period. In a given period, the potential disposable income of a married couple is approximated by estimating the model specified in equation (8), adding a full set of indicators for each calendar year.

The second constraint comes from the specification of the stochastic reapplication process, which determines π_t. We approximate year-to-year transitions to DI by a probit model where the probability of being on DI in year t depends on DI participation in year $t-1$, employment decisions in year $t-1$, and M:

$$
\pi_t(P_{A,t},P_{S,t},D_t,M) \quad = \quad \Phi\left([P_{A,t},P_{S,t},D_t,M]\,\zeta_t\right)
$$

where Φ is the CDF of the standard Normal distribution, M includes a constant, measures of pre-appeal labor market experience and earnings of each spouse, and ζ_t is a vector of parameters for each year t. In each period, the household maximizes expected lifetime utility taken as given its current characteristics, the budget constraint, and the DI transition probabilities. Individuals therefore take account of how working today may lower their chances of being allowed DI in the future (e.g. because it signals work capacity). As a result, D_t becomes a state variable in the dynamic model.

We estimate the parameters of the model in two steps. In the first, we estimate the equation for the transition probabilities, thereby obtaining estimates of ζ_t. In the second step, we jointly estimate the elements of the vector $(\delta_{D,P_A,P_S},\beta_A,\beta_S,\beta_{AS})$ by matching (a) IV estimates of DI participation on I_t,$P_{A,t}$, and $P_{S,t}$ using judge leniency as the instrument; and (b) the covariance matrix between I_t and Q_t for each combination of (D_t,$P_{A,t}$,$P_{S,t}$), where Q_t includes K, M, and dummy variables for education levels of both spouses and the number of children.

We specify and estimate the dynamic model for single and unmarried individuals analogously.