I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15928

# Selective Exercise of Discretion in Disability Insurance Awards

Pilar Garcia-Gomez
Pierre Koning
Owen O'Donnell
Carlos Riumallo Herl

FEBRUARY 2023

DISCUSSION PAPER SERIES

IZA DP No. 15928

# Selective Exercise of Discretion in Disability Insurance Awards

**Pilar Garcia-Gomez**
*Erasmus University Rotterdam and Tinbergen Institute*

**Pierre Koning**
*VU Amsterdam, Tinbergen Institute and IZA*

**Owen O'Donnell**
*Erasmus University Rotterdam and Tinbergen Institute*

**Carlos Riumallo Herl**
*Erasmus University Rotterdam and Tinbergen Institute*

FEBRUARY 2023



ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

# ABSTRACT

## Selective Exercise of Discretion in Disability Insurance Awards*

Variation in assessor stringency in awarding benefits leaves applicants exposed to uninsured risk that could be systematic if discretion were exercised selectively. We test for this using administrative data on applications to the Dutch disability insurance program. We find that discretion is more often exercised in favor of lower-waged applicants. Pre-disability wages drop discontinuously just above disability thresholds for entitlement to partial benefits. Assessors are more likely to discard the highest-paying algorithm-generated job matches that determine earnings capacity and entitlement when evaluating lower-waged applicants. While these applicants benefit on average, they are exposed to greater risk from between assessor variation.

**JEL Classification:** D73, H42, H55

**Keywords:** disability insurance, screening

**Corresponding author:**
Pierre Koning
VU University Amsterdam
Department of Economics, 7A-27
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
E-mail: p.w.c.koning@vu.nl

* Carlos Riumallo Herl received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 840591.

## 1. Introduction

Variation in the exercise of discretion that officials have in adjudicating social insurance and welfare claims leaves applicants exposed to uninsured risk of assignment to a relatively stringent assessor. For example, applicants to US disability insurance (DI) programs face substantial variation in award propensities across claim assessors (Maestas, Mullen, and Strand 2013; French and Song 2014). Less is known about selective exercise of discretion in response to applicants' non-health characteristics. Assessors may show favor to economically vulnerable applicants for whom DI benefits can be particularly valuable (Deshpande and Lockwood 2022).

We test for the selective exercise of discretion in benefit awards using administrative data on the processing and outcomes of applications to the Dutch DI program. This program has a sophisticated claim adjudication process that involves feeding an applicant's functional abilities – determined in a standardized medical examination – education, and basic skills into an algorithm that searches a database to find feasible job matches and calculates the respective earnings capacity (Maestas, Mullen, and Ravesteijn 2021). The proportionate shortfall of earnings capacity from pre-disability earnings defines a *degree of disability* that determines DI entitlement. Within this delineated procedure, there is still scope for assessors to influence awards. They can skip algorithm-generated job matches into the final selection pool or rule that algorithm-generated job matches are infeasible, which would reduce the earnings capacity and raise the benefit entitlement of an applicant. We test for the selective exercise of this discretion across applicants distinguished by pre-disability labor market returns.

We first show indirect evidence of the exercise of discretion. On crossing degree of disability thresholds that determine entitlement to (higher) benefits, there are discontinuous drops in pre-disability wage rates and earnings, and there are upward jumps in a proxy for disability severity. These discontinuities are consistent with assessors exerting relatively more effort to lift lower-waged applicants above the thresholds for receipt of partial benefits and being less lenient with higher-waged applicants. The insurance principle on which the program is founded ensures that for given disability severity and post-disability earnings capacity, benefit entitlement is an increasing function of pre-disability earnings. Consequently, higher-waged applicants can qualify, while more severely disabled lower-waged applicants fail. A sense of unfairness, or an appreciation of the greater value of benefits to the less fortunate, may provoke assessors to intervene to raise the entitlement of lower-waged applicants, while refraining from exerting the same effort in reviewing the claims of higher-waged applicants. We show that assessors interfere to a greater extent with the algorithm-generated job matches for lower-waged applicants.

To gauge the consequences of the selective exercise of discretion, we estimate assessor effects on DI awards that are fixed within applicant pre-disability wage groups but are allowed to differ between them. We find that assessors vary more in deciding the benefits of lower-waged applicants and the leniency of many assessors depends on an applicant's wage. Around a quarter of assessors are more lenient than average with applicants in one wage tercile group while being more stringent than average with applicants in another wage tercile group. Using the distributions of assessor fixed effects, we predict that rejected low-waged (bottom tercile) applicants would have a negligible chance of being awarded benefits if randomly reassigned to another assessor. In contrast, rejected high-waged (top tercile) applicants would have a 13% chance. Opportunities to exercise discretion favorably appear to have been exhausted for rejected low-waged applicants but not for rejected high-waged applicants. Low-waged applicants who were awarded full benefits would face a 7% downside risk of getting lower benefits if randomly reassigned to another assessor, while the respective risk for high-waged applicants is negligeable.

Evidence that US DI awards are prone to misclassification errors (Nagi 1969; Benitez-Silva, Buchinsky, and Rust 2004; Von Wachter, Song, and Manchester 2011; Low and Pistaferri

2015, 2019) and display substantial between-assessor variation (Maestas, Mullen, and Strand 2013; French and Song 2014) has prompted proposals for a more clearly defined adjudication procedure similar to that used in the Netherlands (Maestas 2019; Maestas, Mullen, and Ravesteijn 2021). Limiting discretion afforded to assessors can reduce system noise (Kahneman, Sibony, and Sunstein 2021) and so the element of luck in DI awards that produces horizontal inequity. We show that even in one of the most rule-based DI programs, there is still between assessor variation in DI awards, and there is systematic variation in assessment across applicants.

Evidence of systematic bias in DI awards in relation to a non-health characteristic is rare. Low and Pistaferri (2019) find that US DI false rejection rates are substantially higher for female applicants. This appears to be because women are incorrectly assessed to have higher likelihoods of finding work than men with the same observed health conditions, qualifications, and labor market experience. This is consistent with our hypothesis that Dutch DI assessors favor lower-waged applicants, possibly subconsciously, because they are perceived to have worse labor market opportunities.

Deshpande and Lockwood (2022) demonstrate that, somewhat paradoxically, mismatch between receipt of US DI program benefits and realized health risks is likely to be welfare improving because errors in targeting health risks help to fill gaps in insurance of non-health risks. If lower-waged applicants to the Dutch DI program are exposed to greater uninsured non-health risks, which seems likely at least for labor market risks, then they may benefit more from DI benefits even when disability impacts their earnings potential less than it does for higher-waged applicants. Deshpande and Lockwood (2022) find demand-side evidence of selection on non-health risks that arises through differences in the propensity to apply for DI. We find supply-side evidence of selection through assessors' exercise of discretion in awarding DI.

Between-assessor variation in award rates for DI and other social programs has become a popular instrument to identify effects on labor supply and other outcomes (Doyle 2007; Maestas, Mullen and Strand 2013; Dahl, Kostol, and Mogstad 2014; French and Song 2014; Bakx et al. 2020). In various fields, including law (Kling 2006; Dobbie, Goldin, and Yang 2018; Kleinberg et al. 2018; Bhuller et al. 2020), medicine (Doyle, Ewer, and Wagner 2010; Doyle et al. 2015), and education (Figlio and Lucas 2004), quasi-random assignment to discretion-exercising officials is increasingly exploited for identification – the "judges design". If assignment to an assessor raises the probability of program entry for some applicants but reduces it for others, as we find for low-waged and high-waged DI applications, then monotonicity – strict (Vytlacil 2002), average (Frandsen, Lefgren, and Leslie 2022), and probabilistic (Chan, Gentzkow, and Yu 2022) – is violated and the instrumental variable estimator is not consistent for the local average treatment effect. Chan, Gentzkow, and Yu (2022) demonstrate that monotonicity is violated when medical decision makers have heterogeneous skill in minimizing type I and II errors. Our objective is not to test the monotonicity assumption, but to look beyond the instrumental use of the exercise of discretion to its consequences for the functioning of a disability insurance program.

This paper primarily contributes by being one of the first to deliver evidence of systematic differences in DI award rates. It shows that even in a predominantly rule-based social program, like Dutch DI, claim assessors exercise discretion, and they do so selectively, although not necessarily consciously, to benefit lower-waged applicants.

In the next section, we outline the procedure for assessing applications to the Dutch DI program. Section 3 describes the administrative data used. Section 4 presents evidence first of the selective exercise of discretion in evaluating DI applications and then of its consequences for DI awards. The final section concludes.

## 2. Disability Insurance in the Netherlands

All Dutch employees must contribute to the public DI program. Benefit entitlement is determined by an applicant's *degree of disability*, which is the percentage shortfall of their post-disability earnings capacity from their pre-disability earnings:

$$degree\ of\ disability = \left(1 - \frac{Wage_{post} \times Hours_{post}}{Wage_{pre} \times Hours_{pre}}\right) \times 100, \qquad (1)$$

where $Wage_{pre}$ and $Hours_{pre}$ are, respectively, the hourly wage and the hours worked per week prior to applying for DI, and the potential post-disability wage $(Wage_{post})$ and work hours $(Hours_{post})$ are assessed on the basis of health limitations and educational attainment. For given post-disability earnings capacity, applicants with higher pre-disability earnings have greater degrees of disability and benefit entitlements.

To qualify for any DI benefit, the degree of disability must be at least 35%. Entitlement increases discontinuously at thresholds of 45%, 55%, 65%, and 80%. Applicants above the top threshold are classified as fully disabled and are paid benefits at 70% of pre-disability earnings, up to a maximum of approximately three times the minimum wage. For applicants classified as less than fully disabled, the replacement rate is 70% of the mid-point of the respective degree of disability interval. For example, an applicant assessed as 38% disabled receives 28% (0.7×40%) of pre-disability earnings (Appendix A).

Post-disability earnings capacity $(Wage_{post} \times Hours_{post})$ is estimated following a medical examination by a physician and a subsequent interview with an occupational assessor. The physicians and occupational assessors are employed by the public insurance agency. An examining physician uses a standardized instrument to identify impairments arising from the nature and severity of an applicant's health problem. This results in a functional limitation score that is the sum across 100 items of difficulty in performing work-related tasks (Appendix A).

An occupational assessor feeds this information, along with an applicant's education, basic skills (e.g., driving license and computer skills), and physician-assessed constraints on working hours, into an algorithm that searches the agency's database to identify jobs that the applicant is deemed capable of performing (UWV, 2013).[1] These job matches are ordered from lowest to highest degree of disability (Appendix Figure A1). The algorithm cannot match on functional abilities related to mental capacity and social skills. Hence, the assessor reviews the listed jobs, using information provided by the physician, to determine feasibility. The assessor must review the listed jobs in ascending order of the degree of disability and select the first three that are both feasible and for which there are at least three positions (Figure A2). If possible, the assessor selects some spare jobs (> 3) that imply higher degrees of disability. If it is not possible to select three jobs, then the applicant is classified as fully disabled with 100% degree of disability. Otherwise, the applicant's potential wage ($Wage_{post}$) is the median wage of the three jobs with the lowest degrees of disability, and potential hours ($Hours_{post}$) are the minimum hours across these three jobs.

If the assessor takes no further action, the resulting degree of disability is used to calculate benefit entitlement. However, the assessor may exclude certain job matches, switching to spares, and re-calculating the degree of disability. This would lead to upward revision of the degree of disability. This process can continue interactively, giving the occupational assessor the opportunity to observe the impact of different job selections on benefit entitlement. The occupational assessor can also ask the physician to reconsider whether a functional limitation score is a complete and accurate reflection of the applicant's capacity. Occasionally, the

1 For each job, the database contains the wage and hours, as well as physical and mental capacities required to perform job-related tasks. The insurance agency collects this information by visiting employers throughout the country. Each visit focusses on starting positions that could potentially have been filled by people with functional impairments. These need not be vacant positions.

physician makes changes that facilitate matching of functional impairments to (other) jobs and this reduces the degree of disability (Appendix A).

While the assessment process is reasonably tightly defined, it leaves scope for the exercise of discretion. Occupational assessors have latitude to raise benefit entitlement by ruling some job matches infeasible in the first matching list or from the final job selection. Discretion could be used indiscriminately or selectively. The fact that lower-waged applicants must be more functionally impaired to reach a given benefit entitlement (Appendix Figure C1) may motivate some assessors to intervene with job selections. Consider applicant *L* who is deemed to have more functional limitations than applicant *H*, who has a higher pre-disability wage. The wage difference may dominate, such that, even if both applicants were assessed as capable of performing the same jobs, *H* would have a higher degree of disability because of a greater loss of earnings capacity. Applicant *H* may qualify for (higher) DI benefit, while *L* would not. Potentially, an assessor could respond by classifying better-paying job matches found by the algorithm as infeasible for *L*, which would raise the degree of disability to an extent sufficient for *L* to qualify for (higher) DI benefit. Or the assessor may exert greater effort in finding better-paying job matches for *H,* which would reduce the degree of disability and benefit entitlement of *H*.

Assessors need not consciously compare applicants in this way. They may simply learn to identify the type of applicants who qualify for (higher) DI and subconsciously exert greater effort to help those who otherwise would not. Variation in the exercise of discretion across occupational assessors need not reflect differences in consciously motivated actions. It could reflect differences in daily routines or informal guidelines issued by managers.

## 3. Data

We use administrative records covering the universe of DI applications from January 2006 to July 2017 (Appendix B). We exclude applications that were decided without using the algorithm to match applicant characteristics with feasible jobs. These are cases for which the medical examination was sufficient to establish full disability. For each of the remaining applications (n≈400,000), we have data on a) demographics, education, employment contract (permanent or temporary), industry, wage rate $(Wage_{pre})$, and work hours $(Hours_{pre})$ at the time of application, b) diagnosis and functional limitations score from the medical examination, and c) potential wage rate $(Wage_{post})$, potential work hours $(Hours_{post})$, and degree of disability from the occupational assessment. For each application, we also observe the number of medical examinations and occupational assessments conducted, as well as job characteristics (hours and hourly wages) in the final set of jobs used to calculate the degree of disability (Figure A2). For each application from 2011 to 2017, anonymized codes uniquely identify the examining physician, the occupational assessor, and the assessment office.

The average age of DI applicants is 46 years (Appendix Table C1). A slight majority is female (52%). Around a quarter (26%) of applicants have no more than compulsory education. Before application, a slight majority (52%) was either on a temporary contract or unemployed, average work hours were about 32 per week, and the average hourly wage was 16.53 euros (2010 prices). The most prevalent main diagnosis is a musculoskeletal condition (35%), followed by a mental health problem (30%). The average functional limitations score indicates difficulty in almost 13 (out of 100) domains of work-related tasks. The average degree of disability (37.5%) is just above the 35% threshold at which an applicant is entitled to the lowest partial DI benefit. Around 59% of applicants do not reach this threshold, while 21% have a degree of disability of at least 80% and qualify for the full benefit.

## 4. Results

### *4.1 Exercise of discretion - evidence*

Degree of disability is positively related to pre-disability earnings and functional limitations score.[2] If discretion were not exercised selectively, each of these determinants would be expected to increase smoothly with the degree of disability through the DI entitlement thresholds. We test this by estimating these regressions,

$$Y_i = \beta_\tau \cdot 1(DD_i \geq T_\tau) + f_{\tau L}(DD_i) \cdot 1(DD_i < \mathrm{T}_\tau) + f_{\tau R}(DD_i) \cdot 1(DD_i \geq \mathrm{T}_\tau) + \varepsilon_{\tau i}\,, \tag{2}$$

where $Y_i$ is, in turn, the pre-disability hourly wage, hours worked per week, weekly earnings or the functional limitations score of applicant *i*, $DD_i$ is the degree of disability, $T_\tau$ is a threshold equal to 35%, 45%, 55%, 65%, or 80% in the respective regression, 1(.) is the indicator function, and $f_{\tau L}(DD_i)$ and $f_{\tau R}(DD_i)$ are flexible functions of the degree of disability, which are specified as third order polynomials in the main estimates. We use the same coverage error optimal bandwidth on each side of the threshold and conduct robust biased-corrected inference (Calonico et al. 2020).

Consistent with the selective exercise of discretion, Table 1 and Figure 1 show discontinuities at entitlement thresholds. The pre-disability wage drops significantly ($p < 0.01$) on crossing each of the three lower thresholds from below. The estimated €1.37 drop at the 35% threshold implies that applicants given a degree of disability just sufficient to get the lowest benefit had a wage that was 7.0% lower than the wage of those just denied DI. The direction of this effect runs counter to the built-in propensity for the degree of disability to increase with pre-disability earnings. The wage discontinuities are greater – both in absolute

[2] Pre-disability earnings enter the calculation of the degree of disability directly, eq. (1). The influence of the functional limitations score is indirect. It operates through matching of abilities to jobs in the insurance agency's database and, consequently, determination of post-disability earnings capacity.

and relative terms – at the 45% and 55% thresholds where the benefit increases. At the 65% thresholds, the estimated wage drop is smaller and not significant ($p \geq 0.1$). At the top threshold that determines entitlement to full benefit, the point estimate is larger and the lack of significance could possibly be due to the smaller sample size. There are significant ($p < 0.05$) discontinuous downward shifts in pre-disability weekly earnings at the 35%, 45%, and 55% thresholds that are driven by drops in wages rather than hours (see also Figure C2). The point estimate of the earnings effect at the 80% threshold is larger in magnitude than the significant estimates at the two lowest thresholds.

There are significant ($p < 0.05$) upward discontinuities in the functional limitations score as the degree of disability crosses the 35% and 45% thresholds from below. Applicants given a degree of disability just sufficient to qualify for DI (≥ 35%) were assessed to have 1.66 more functional limitations than those who just failed to qualify, which corresponds to a 14.1% increase. The point estimates indicate upward jumps in limitations at the 55%, 65%, and 80% thresholds, although none of these is close to significance.

TABLE 1. DISCONTINUITIES IN DETERMINANTS OF DISABILITY INSURANCE AWARDS AT ENTITLEMENT THRESHOLDS

| | Degree of disability threshold for DI entitlement | | | | |
|---|---|---|---|---|---|
| | 35% | 45% | 55% | 65% | 80% |
| Wage per hour (€) | -1.365 | -1.637 | -2.197 | -0.644 | -1.820 |
| | (0.226) | (0.296) | (0.499) | (0.544) | (1.195) |
| Hours per week | 0.705 | 0.972 | -0.122 | -0.058 | 0.365 |
| | (0.352) | (0.368) | (0.381) | (0.359) | (0.591) |
| Earnings per week (€) | -30.666 | -37.586 | -86.993 | -22.528 | -46.966 |
| | (9.766) | (13.463) | (18.201) | (20.602) | (47.144) |
| Functional limitations score | 1.655 | 0.984 | 0.608 | 0.131 | 0.697 |
| | (0.306) | (0.387) | (0.418) | (0.384) | (0.751) |
| N | 40152 | 20081 | 17529 | 20736 | 6426 |

*Notes*: Bias-corrected robust regression discontinuity estimates of $\beta_\tau$ from model (2) with $f_{\tau L}(DD_i)$ and $f_{\tau R}(DD_i)$ each specified as a different third-order polynomial and the same (coverage error probability) optimal bandwidth on each side of the threshold. Robust standard errors in parentheses. Wage, hours, and earnings are prior to application for DI.

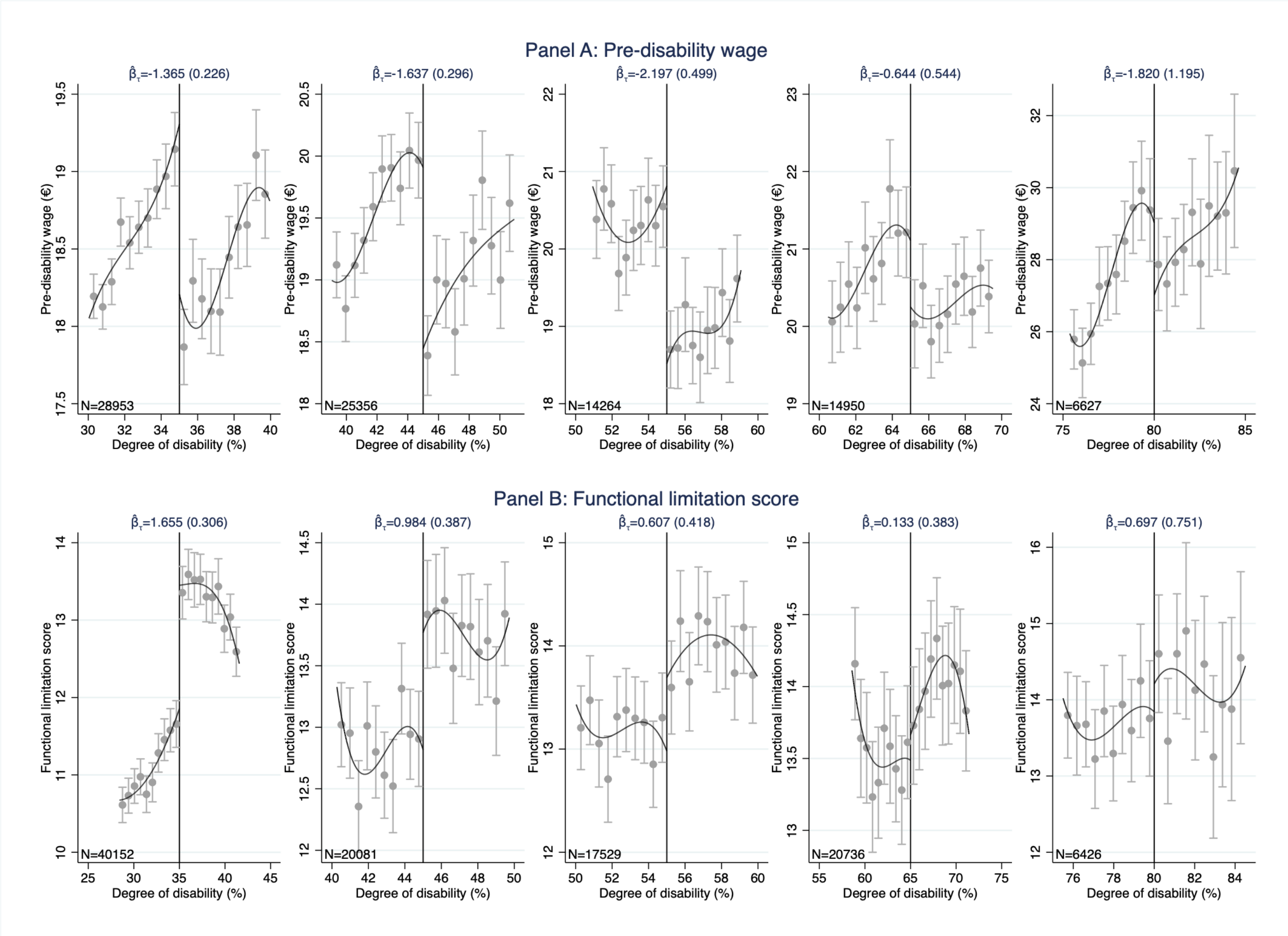


FIGURE 1. DISCONTINUITIES IN PRE-DISABILITY WAGE AND FUNCTIONAL LIMITATIONS AT DISABILITY INSURANCE ENTITLEMENT THRESHOLDS

*Note*: Bias-corrected robust regression discontinuity with a different third order polynomial on each side of the respective threshold and a threshold-specific optimal coverage error bandwidth. Dots represent bin averages (10 bins on each side of threshold). Interval lines show 95% confidence intervals.

The estimates of threshold discontinuities in the determinants of DI entitlement are robust to using different functions of the degree of disability and bandwidths (Table C2). There are some differences in magnitudes, but the direction and significance of the discontinuities do not change.

One may hypothesize that these discontinuities arise from lumpiness in the wage distribution of the algorithm-generated job matches or some other source of nonlinearity that may exist even without selective exercise of discretion and is not fully captured by the flexible functions specified on either side of the respective threshold. If this were the case, then discontinuities would emerge also at degrees of disability that are not entitlement thresholds. To test this, we estimate equation (2) for the pre-disability wage and the functional limitations score at degree of disability $T_\tau$ set equal to every second percentage point from 21% to 71% and 74% to 84%. We find only two (out of 54) significant ($p < 0.05$) discontinuities that are not at DI entitlement thresholds (Figure C3). This predominance of null placebo effects suggests that the three significant wage discontinuities and the two significant limitations score discontinuities that are found across only five entitlement thresholds are unlikely to be attributable to unmodelled nonlinearities that are built into the determination of degree of disability. They are more likely to arise from selective exercise of discretion close to the thresholds.

The discontinuities at each threshold could result from lower-waged applicants in worse health being hoisted up to reach that threshold or from higher-waged applicants with fewer limitations being held below it. Consistent with the latter mechanism, Figure 2 shows that in the top wage tercile and the bottom two functional limitations terciles, applications are bunched below degree of disability thresholds, leaving missing mass in regions where entitlement increases. Non-parametric bunching estimates (Cattaneo et al. 2021) confirm discontinuities in the distributions of high-waged and lower-limitations applications at the three lowest thresholds (Table C3). In the bottom two terciles of the wage distribution, there is either no significant bunching or an excess mass of applications at the three lowest thresholds. There is no bunching the top of the limitations distribution.[3]

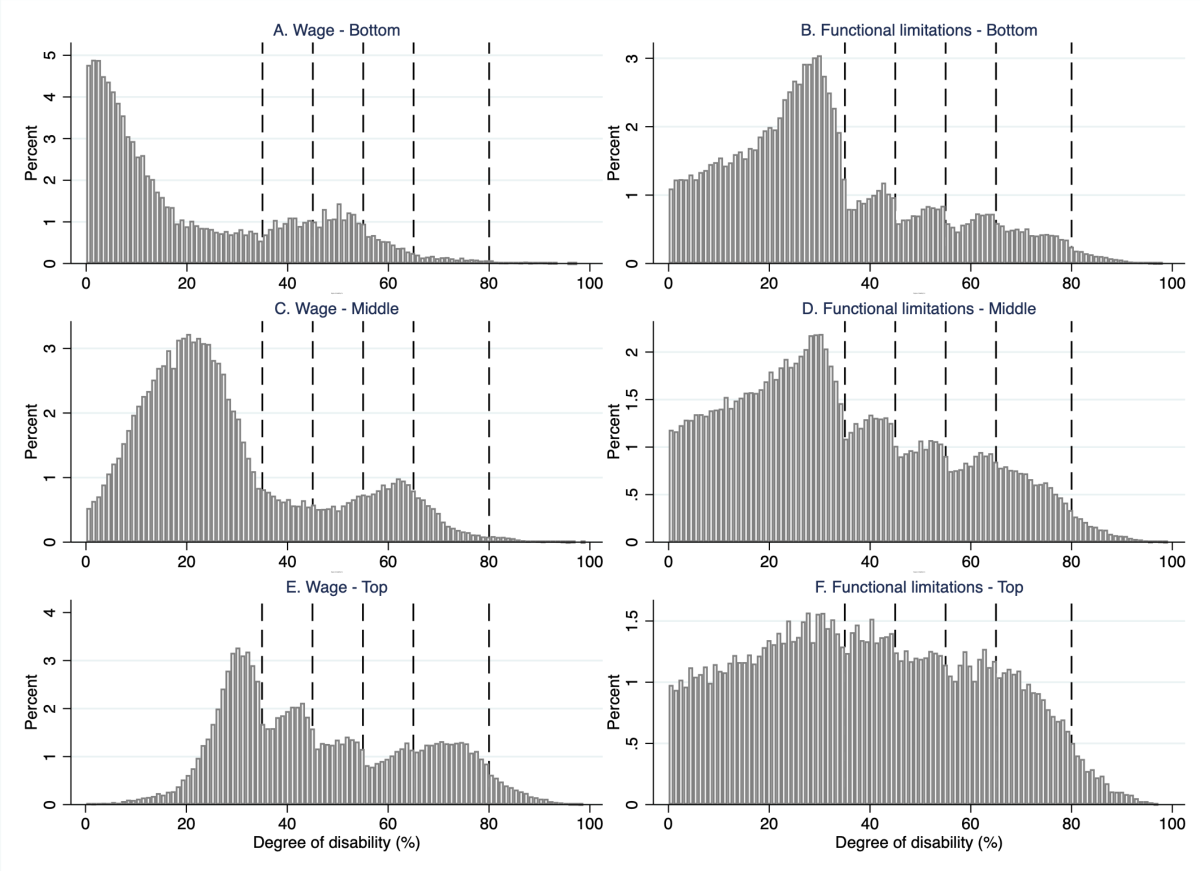


FIGURE 2. DI APPLICATIONS BY DEGREE OF DISABILITY STRATIFIED BY WAGE AND FUNCTIONAL LIMITATIONS SCORE TERCILE GROUPS

*Notes*: Each panel shows a distribution of DI applications by degree of disability. In panels A, C, and E applications are stratified into the bottom, middle and top third of the pre-disability hourly wage distribution, respectively. In panels B, D, and F applications are stratified into the bottom, middle, and top third of the functional limitation score distribution, respectively. To focus on thresholds critical to benefit entitlement, in this figure we exclude applications with degree of disability of 0 (~30% of applications) and 100 (~20% of applications). Sample sizes: 35,008, 78,542, and 78,825 for panels A, C, and E, respectively; 53,164, 97,559, and 41,815 for panels B, D, and F, respectively. The number of applications is not the same across groups because the terciles are calculated including those with degree of disability of 0 and 100. Dashed lines represent the DI entitlement thresholds. See Table C3 for bunching estimates and Figure C4 for the unstratified density of all applications.

Less exercise of discretionary effort in favor of high-waged applicants is also indicated by the fact that discontinuities in the pre-disability wage and the functional limitations score at the

[3] There is bunching below the three lowest degree of disability thresholds in the unstratified distribution of all applications (Figure C4 and Table C3).

35% threshold are evident and significant only among applicants in the top tercile of the wage distribution (Figure C5). A larger upward jump in functional limitations is needed to lift these applicants over the qualification threshold (Figure C5).

Analyses reported in Appendix D provide further evidence that the discontinuities in pre-disability wages at entitlement thresholds are likely due to occupational assessors selectively exercising discretion across the wage distribution. At the 35% threshold, there is an upward shift in the proportion selected job matches that imply a degree of disability sufficient to qualify for DI that is substantially larger for lower-waged applicants (Table D2). This is consistent with assessors exerting greater effort to secure DI entitlement for these applicants. At each threshold, low-waged applicants have a larger difference between the degree of disability finally used to determine their DI entitlement and the degree of disability that would have resulted from strictly adhering to the guideline of using the three highest-paying algorithm-generated job matches (Tables D3 & D4). For example, in the vicinity of the 35% threshold, by ruling higher-paying job matches infeasible, assessors raise the degree of disability of applicants in the bottom third of the wage distribution by 7.5 percentage points (pp), on average. For applicants in the top third of the wage distribution, the respective average increase in the degree of disability is only 0.86 pp. Consistently, there is a higher probability that lower-waged applicants would have received lower (or no) benefits if the three highest-paying jobs had been used to calculate their degree of disability (Table D3 & D4). For around 60% of assessors, the likelihood of intervening by using jobs other than the three highest-paying matches is higher when the applicant is low-waged rather than high-waged (Figure D2).

### *4.2 Exercise of discretion - consequences*

To gauge the importance of the exercise of discretion and of systematic differences in its prevalence, we estimate assessor fixed effects (FE) in determining the degree of disability and use them to simulate DI awards under counterfactual assignments of applicants to assessors.

To allow assessor behavior to depend on applicants' pre-disability wages, we stratify on that variable and estimate,

$$DD_{ijrt}^{g} = \alpha^{g} + \lambda_{j}^{g} + \boldsymbol{X}_{i}\boldsymbol{\beta}^{g} + \delta_{r}^{g} + \tau_{t}^{g} + \varepsilon_{ijrt}^{g}\,, \tag{3}$$

where $DD_{ijrt}^{g}$ is the final degree of disability of applicant $i$ in wage tercile group $g \in \{low,\ middle,\ high\}$ and ZIP Code area *r* who is assigned to assessor *j* in year *t*. $\lambda_{j}^{g}$ are wage group specific assessor FE, $\boldsymbol{X}_{i}$ is a vector of controls that includes the applicant's age, sex, employment contract at the time of application, functional limitations score, limitations on daily or weekly work schedule, main diagnosis, educational attainment, and hours worked at application. $\delta_{r}^{g}$ and $\tau_{t}^{g}$ are ZIP Code and year FE, respectively, and $\varepsilon_{ijrt}^{g}$ is a stochastic error. We estimate these regressions by OLS using applications assigned to assessors who handled at least 10 applications per wage group and 50 in total.[4]

Figure 3 shows a scatter plot of mean-centered assessor FE estimated using applicants in the highest wage group against the respective FE obtained from the lowest wage groups. Greater horizontal than vertical dispersion of the effects indicates greater variation between assessors in determining the degree of disability of low-waged applicants.[5] Assignment to an assessor carries more risk for a low-waged applicant. There is a moderate correlation ($\hat{\rho} = 0.357$) between the assessor FE for low-waged and high-waged applicants. Around a quarter (24.7%) of the assessors are more lenient (higher degree of disability) than average when evaluating applicants from one wage group while being more stringent than average when evaluating applicants from the other group. This discordance is not the mechanical result of compensatory

[4] On average, an assessor handled 116 applications (Table C4). About half (50.4%) of the 2,635 assessors meet the inclusion criteria of handling at least 10 applications within each wage tercile group and at least 50 applications in total. Most of the excluded assessors handled less than 10 applications in total (Figure C6).

[5] See Figure C7 for clearer graphical evidence of the greater variance of the FE in assessments of lower-waged applicants and Table C5 for confirmation that the standard deviation and the inter-quartile range of the FE are both larger for lower-wage groups.

decisions to remain within a cap on the proportion of applications an assessor can award because there is no such cap.

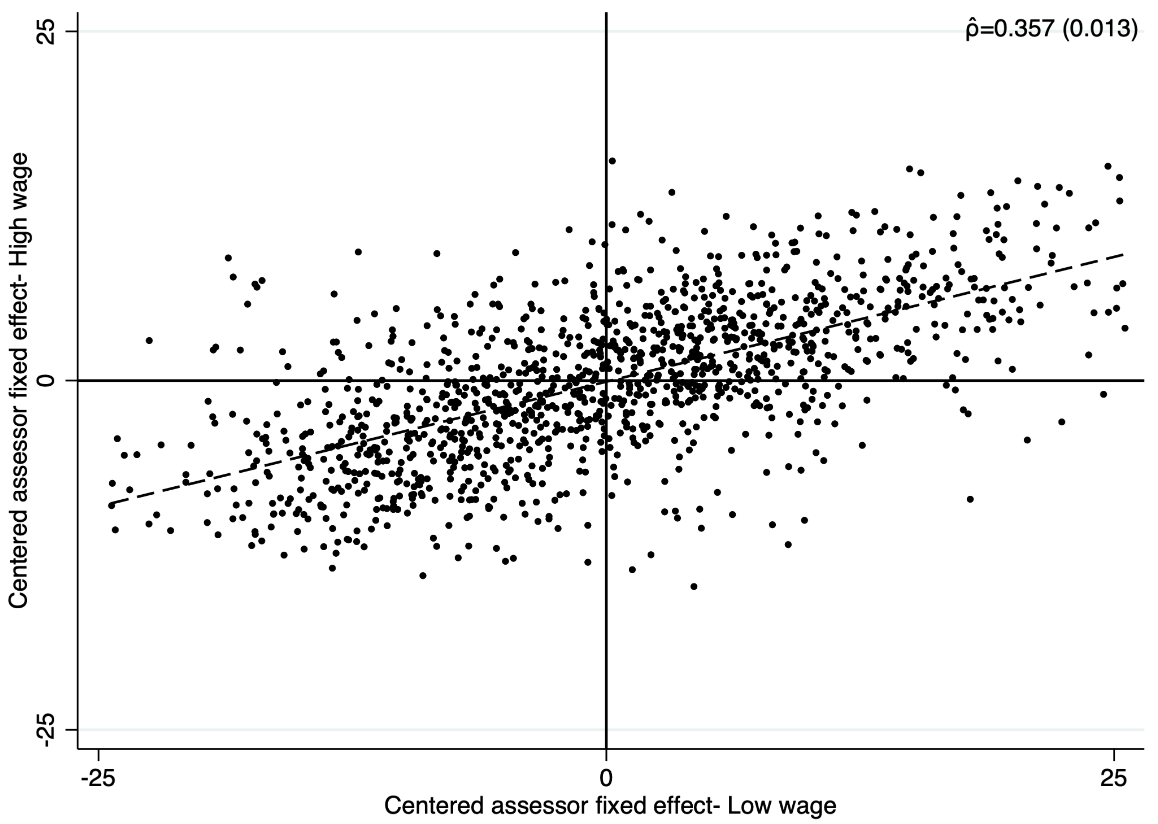


FIGURE 3. SCATTER OF ASSESSOR FIXED EFFECTS FOR HIGH- AND LOW-WAGED APPLICANTS

*Notes*: Each dot represents one assessor. Estimates of fixed effects are obtained from eq. (3) and then centered at the respective wage tercile group mean. Positive values indicate assessors who are more lenient than average. $\hat{\rho}$ is the estimated correlation coefficient, with the standard error in parentheses.

We predict each applicant's degree of disability if assigned to a different assessor by substituting the estimated FE of that assessor $(\hat{\lambda}_k^g)$ for the estimated FE of the assessor to which the applicant is actually assigned $(\hat{\lambda}_j^g)$,

$$\widetilde{DD}_{ikrt}^g = DD_{ijrt}^g + \left(\hat{\lambda}_k^g - \hat{\lambda}_j^g\right) \quad k \neq j. \tag{4}$$

Repeated prediction under assignment to each assessor gives a distribution of counterfactual degrees of disability for each applicant. We use this to calculate the probabilities of being awarded no benefit $\left(Pr\left(\widetilde{DD}_{ikrt}^g < 35\right)\right)$, partial benefit $\left(Pr\left(35 \leq \widetilde{DD}_{ikrt}^g < 80\right)\right)$, and full benefit $\left(Pr\left(\widetilde{DD}_{ikrt}^{g} \geq 80\right)\right)$ if randomly assigned to an assessor. We average these probabilities within each wage tercile group crossed with actual DI award category (no benefit, partial benefit, and full benefit) and show the results in panel A of Table 2.[6]

TABLE 2. ACTUAL AND COUNTERFACTUAL DISABILITY INSURANCE AWARDS BY WAGE TERCILE

| | Counterfactual Award<br>No benefit<br>(DD<35%)<br>% | Partial benefit<br>(DD=35-79%)<br>% | Full benefit<br>(DD≥80%)<br>% | Number of applicants |
|---|---|---|---|---|
| *Panel A) Random assignment to an assessor* | | | | |
| Actual Award | | | | |
| No benefit | | | | |
| Low wage | 100.00 | 0.00 | 0.00 | 61,260 |
| Middle wage | 97.70 | 2.30 | 0.00 | 54,429 |
| High wage | 87.05 | 12.95 | 0.00 | 25,868 |
| Partial benefit | | | | |
| Low wage | 7.15 | 92.85 | 0.00 | 9,147 |
| Middle wage | 0.00 | 100.00 | 0.00 | 18,863 |
| High wage | 0.00 | 100.00 | 0.00 | 43,556 |
| Full benefit | | | | |
| Low wage | 0.00 | 7.07 | 92.93 | 25,508 |
| Middle wage | 0.00 | 1.11 | 98.89 | 24,029 |
| High wage | 0.00 | 0.00 | 100.00 | 25,530 |
| *Panel B) All assigned to most lenient assessor* | | | | |
| Actual Award | | | | |
| No benefit | | | | |
| Low wage | 67.08 | 32.91 | 0.01 | 61,260 |
| Middle wage | 32.64 | 67.36 | 0.00 | 54,429 |
| High wage | 9.77 | 90.23 | 0.00 | 25,868 |
| Partial benefit | | | | |
| Low wage | 0.00 | 66.50 | 33.50 | 9,147 |
| Middle wage | 0.00 | 62.84 | 37.16 | 18,863 |
| High wage | 0.00 | 68.43 | 31.57 | 43,556 |
| *Panel C) All assigned to most stringent assessor* | | | | |
| Actual Awards | | | | |
| Partial benefit | | | | |
| Low wage | 78.29 | 21.71 | 0.00 | 9,147 |
| Middle wage | 46.31 | 53.69 | 0.00 | 18,863 |
| High wage | 39.40 | 60.60 | 0.00 | 43,556 |
| Full benefit | | | | |
| Low wage | 0.00 | 86.44 | 13.56 | 25,508 |
| Middle wage | 0.00 | 66.59 | 33.41 | 24,029 |
| High wage | 0.00 | 30.78 | 69.22 | 25,530 |

Notes. In panel A, each cell gives the average percent chance of a degree of disability (DD) in the interval indicated by the respective column heading. The percent chance is obtained for each applicant from the cumulative distribution of $\widetilde{DD}_{ikrt}^{g}$, which is calculated using eq. (4) for all assessors ($k$). The individual level percent chances are then averaged over all applicants in the same actual DI award category and wage tercile group indicated by the row heading. In panel B, each cell gives the row percent of applicants with $\widetilde{DD}_{ikrt}^{g}$ calculated from eq. (4) with $\hat{\lambda}_{k}^{g} = max\left(\hat{\lambda}_{1}^{g}, \dots, \hat{\lambda}_{K}^{g}\right)$ in the interval given by the respective column heading. Panel C cells are constructed analogously with $\hat{\lambda}_{k}^{g} = min\left(\hat{\lambda}_{1}^{g}, \dots, \hat{\lambda}_{K}^{g}\right)$. Low wage, middle wage, and high wage are applicants in the bottom, middle, and top third of the pre-disability wage distribution, respectively. DD = degree of disability. The extreme right-hand column gives the number of applicants in each DI category-wage group.

[6] Figure C8 gives cumulative distributions of counterfactual degrees of disability – $Pr\left(\widetilde{DD}_{ikrt}^{g} < X\right)$ at each $X \in \{1, \dots, 100\}$ – averaged within each wage tercile group crossed with actual DI award category.

Low-waged applicants who were denied DI (DD < 35) had a negligible chance (<0.01%) of being awarded partial benefit if each had been assigned to a different assessor. When assessing these applicants, assessors appear to have exhausted opportunities to exercise discretion by ruling higher-paying job matches infeasible. This is not true for rejected high-waged applicants. On average, they would have a 13 percent chance of getting partial benefit if assigned to another assessor.

For those actually paid partial benefits, only the low-waged applicants are effectively at risk of an outcome other than a different level of partial benefits if randomly assigned to another assessor. For this group, the downside risk of losing entitlement (7.3%) is substantially greater than the negligible upside risk of acquiring full benefits. Similarly, among those awarded full benefits, around 7% of the low-waged applicants would have received only partial benefits if assigned to another assessor, compared with a negligeable risk in the high-wage group. The fact that benefits paid to accepted low-waged applicants are more contingent on the assessor to which they are assigned is consistent with some assessors being more likely to rule higher-paying job matches as infeasible for this group.

The bottom two panels of Table 2 give, for each wage and actual DI award category, the predicted percentages of applicants who would be awarded no benefit, partial benefit, and full benefit if all were assigned to the most lenient assessor within the respective wage group $\left(\hat{\lambda}_k^g = max(\hat{\lambda}_{1,}^g \dots, \hat{\lambda}_K^g), \text{ panel B}\right)$ and to the most stringent assessor $\left(\hat{\lambda}_k^g = min(\hat{\lambda}_1^g, \dots, \hat{\lambda}_K^g), \text{ panel C}\right)$. Since there is only a small probability that an applicant would be assigned to the most extreme assessor in each direction, these counterfactuals overstate the consequences of the exercise of discretion. We show them to demonstrate that our finding of differential effects across wage groups is robust to restricting attention to extreme scenarios.[7]

If all applicants were assigned to the most lenient (within wage group) assessor, then around 33% of rejected low-waged applicants and 90% of rejected high-waged applicants would get partial benefits. This implies that the fraction of *never takers* among rejected applicants is much larger for the low-waged than it is for the high-waged. In contrast, assignment to the most stringent assessor would be most consequential for the lower-wage groups. Among those awarded partial benefits, assignment to the most stringent assessor would have resulted in rejection of around 79% of the low-wage applications but only around 39% of the high-wage applications. About 87% of low-waged applicants, but only 31% of high-waged applicants, who were awarded full benefits would have received partial benefits if assigned to the most stringent assessor.

The downside risk of losing entitlement if assigned either randomly to another assessor or specifically to the most stringent assessor is larger for low-waged applicants. The upside risk of gaining entitlement if reassigned either randomly or specifically to the most lenient assessor is larger for high-waged applicants. Both patterns are consistent with selective exercise of discretion in favor of the low-wage group.

## 5. Conclusions

As with many other social programs, disability insurance awards are decided by applying objective rules in combination with judgement to take account of characteristics and circumstances that are difficult to measure and codify. Giving assessors discretion makes use

[7] For their main estimate of the fraction of US DI applicants whose awards *could* be changed by assessor assignment, Maestas et al. (2013) use the difference in award probability that would arise from the full range of difference in assessor FEs (from most stringent to most lenient). They acknowledge (their footnote 43) an order of magnitude smaller estimate of the fraction whose DI awards *would* be changed by eliminating assessor variation in award rates.

of additional information they can glean from applicants at the inevitable cost of inconsistency and horizontal inequity. Our analysis of applications to the Dutch DI program reveals that judgements do not only generate random between-assessor variation in award propensities – system noise (Kahneman, Sibony, and Sunstein 2021) – but can also produce systematic differences in awards across applicants distinguished by labor market characteristics. We find that discretion is more likely to be exercised in favor of lower-waged applicants. Discontinuous drops in pre-disability wages just above entitlement thresholds reflect upward jumps in the fraction of lower-waged applicants who just qualify for higher benefits. This is contrary to what would be expected to arise if rules that make entitlement an increasing function of the pre-disability wage were applied strictly. It occurs because some assessors are less likely to follow a guideline to calculate earnings capacity and entitlement using the highest-paying algorithm-generated job matches when they are evaluating lower-waged applicants. They exert effort to rule out job matches for these applicants and are less likely to do this for higher-waged applicants. While this selective exercise of discretion benefits lower-waged applicants on average, they are exposed to greater uninsured risk due to larger between assessor variation in awards.

Selective exercise of discretion in favor of lower-waged DI applicants may arise from claim assessors – the street-level bureaucrats (Lipsky 2010) who implement disability insurance policy – acting with a sense of fairness that conflicts with the insurance principle of the DI program. If lower-waged applicants value DI benefits more, possibly because they are exposed to more uninsured non-health risks related to their disadvantaged labor market position (Deshpande and Lockwood, 2022), then the selective discretion exercised by assessors could increase the social value of the program, although we have not tested that hypothesis. While the relatively high social safety net in the Netherlands would be expected to limit the scope for DI to cover otherwise uninsured non-health risks, denial of program benefits to lower-waged applicants can still result in substantial relative income losses, particularly for those without working partners, and costly job search obligations.

## References

**Bakx, P., B. Wouterse, E. van Doorslaer and A. Wong.** 2020. "Better off at home? Effects of nursing home eligibility on costs, hospitalizations and survival." *Journal of Health Economics*, 73, 102354.

**Benitez-Silva, H., M. Buchinsky and J. Rust.** 2004. "How Large Are the Classification Errors in the Social Security Disability Award Process?," *NBER Working Paper 10219*, Cambridge, MA: National Bureau of Economic Research.

**Bhuller, M., G.B. Dahl, K.V. Loken, and M. Mogstad.** 2020. "Incarceration, recidivism, and employment." *Journal of Political Economy*, 128, 1269–1324

**Calonico, S., M.D. Cattaneo, and M.H. Farrell. 2020.** "Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs." *The Econometrics Journal*, 23(2), 192–210.

**Cattaneo, M., M. Jansson and X. Ma.** 2021. "Local Regression Distribution Estimators." *Journal of Econometrics* available online 6 March 2021 https://doi.org/10.1016/j.jeconom.2021.01.006

**Chan, D.C., M. Gentzkow, and C. Yu.** 2022. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." *Quarterly Journal of Economics*, 137(2), 729–783

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

**Dobbie, W., J. Goldin, and C.S. Yang**. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, 108, 201–240

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

**Doyle, J.J., S.M. Ewer, and T.H. Wagner**. 2010. "Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams." *Journal of Health Economics*, 29, 866–882.

**Doyle, J.J., J.A. Graves, J. Gruber, and S. Kleiner**. 2015. "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns." *Journal of Political Economy*, 123, 170–214

**Figlio, D.N. and M.E. Lucas**. 2004., "Do High Grading Standards Affect Student Performance?" *Journal of Public Economics*, 88, 1815–1834

**Frandsen, B.R., L.J. Lefgren, and E.C. Leslie**. 2022. "Judging Judge Fixed Effects," *American Economic Review* forthcoming.

**French, E. and J. Song.** 2014. "The Effect of Disability Insurance Receipt on Labor Supply." *American economic Journal: economic policy*, 6(2), 291-337.

**Kahneman, D., O. Sibony, and C.R. Sunstein**. 2021. *Noice: A Flaw in Human Judgement*. London: William Collins.

**Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan**. 2018. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics*, 133, 237–293.

**Kling, J.R.** 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review*, 96, 863–876.

**Lipsky, M.** 2010. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service*. New York: Russell Sage Foundation.

**Low, H. and L. Pistaferri**. 2015. “Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off.” *American Economic Review* 105, 2986-3029.

**Low, H. and L. Pistaferri.** 2019. "Disability Insurance: Error Rates and Gender Differences," *NBER Working Paper 26513*, Cambridge, MA: National Bureau of Economic Research.

**Maestas, N**. 2019. "Identifying Work Capacity and Promoting Work: A Strategy for Modernizing the SSDI Program." *Annals of the American Academy of Political and Social Science*, 686(1), 93-120.

**Maestas, N., K. Mullen and A. Strand.** 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of Ssdi Receipt." *American Economic Review*, 103(5), 1797-829.

**Maestas, N., K. Mullen and B. Ravesteijn.** 2021. “*Applying Aspects of Disability Determination Methods From the Netherlands in the U.S.*” NBER Retirement and Disability Research Center Paper NB21-08. Cambridge, MA: National Bureau of Economic Research.

**Nagi, S**. (1969). *Disability and Rehabilitation*. Columbus, OH: Ohio State University Press

**UWV.** 2013. "Basisinformatie Cbbs," UWV, Amsterdam: UWV.

**Von Wachter, T., J. Song, and J. Manchester**. 2011. “Trends in Employment and Earnings of Allowed and Rejected Applicants to the Social Security Disability Insurance Program.” *American Economic Review*, 101(7): 3308–29.

**Vytlacil, E**. 2002. “Independence, Monotonicity, and Latent Index Models: An Equivalence Result.” *Econometrica*, 70, 331–341.

Supplementary Material

# Selective Exercise of Discretion in Disability Insurance Awards

PILAR GARCIA-GOMEZ, PIERRE KONING, OWEN O'DONNELL, CARLOS RIUMALLO HERL*



* Garcia-Gomez: Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Netherlands, Tinbergen Institute (email: garciagomez@ese.eur.nl); Koning: School of Business and Economics, VU Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands, Tinbergen Institute and IZA (email: p.w.c.koning@vu.nl); O'Donnell: Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Netherlands, Tinbergen Institute (email: odonnell@ese.eur.nl); Riumallo Herl: Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Netherlands, Tinbergen Institute (email: riumalloherl@ese.eur.nl). Carlos Riumallo Herl received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 840591.

## 1. Appendix A. Institutional details of Dutch disability insurance program

In this section, we present in more detail the organization of the Dutch Disability insurance program. We describe how the benefit amounts are defined, the medical and occupational evaluation, as well as the steps taken to define the remaining work capacity.

*Benefit amount*. Depending on employment history, disability insurance (DI) benefits equal 70% of pre-disability earnings for a period of up to 24 months. Thereafter, benefits continued to be linked to pre-disability earnings only if actual earnings exceed 50% of potential earnings capacity. If they do not, then benefits are tied to the statutory minimum wage instead (Koning and Van Sonsbeek, 2017).

*Medical examination*. The medical assessor uses a standardized instrument to make a detailed list of physical, social, and mental tasks an applicant would have difficulty performing. Examples of tasks are being able to work independently, to hear in certain environments, and to use hands and fingers. This list is transformed into a profile of the applicant's functional ability to perform work-related tasks. For example, if the applicant is not able to lift an object of a specified weight without difficulty, then they will be deemed incapable of performing jobs that involve lifting. The physician also determines whether the applicant could workday or night shifts and the maximum number of work hours per day and workdays per week. The physician may test the applicant's ability to perform certain tasks, but the assessment is mainly based on questions to the applicant about daily routines and functional limitations, and on information supplied by occupational and treating physicians.

*Occupational assessment*. The employment agency's database distinguishes between types of jobs, jobs, and positions. A *job type* is a cluster of jobs within a standardized occupational classification. To calculate potential earnings capacity, and so the corresponding degree of disability, at least three feasible job types must be found. A *job* within a job type may have multiple *positions* that are used to calculate the median wage per job type. For example, *domestic worker* is a job type, a *domestic services employee at Mr. Clean company* is a job, and if Mr. Clean has three such employees, then there are three positions. A job type can be used to calculate potential earnings if there are at least three positions from one or more jobs within that type. The database includes job types that are available in each region, although there are some differences across regions.

At first, the occupational assessor matches information on an applicant's physical functional limitations, education, work hours capacity, and basic skills (e.g., driving license, computer literacy) with the jobs database to identify jobs the applicant is potentially capable of performing. The database holds information on the wage and hours for each job, as well as the physical and mental functioning required to perform the job's tasks.

This results in a list of potentially feasible jobs ordered (from lowest to highest) by degree of disability that is inversely related to the potential earnings in each job (Figure A1). The assessor reviews these jobs for actual suitability since the algorithm does not match automatically on all functional abilities, particularly those related to mental and social capacity. For this, the occupational assessor can use the functional limitations score provided by the physician. The occupational assessor must review the listed jobs in ascending order from lowest to highest degree of disability (and so, benefit entitlement).

**Client dossier: Mr. D.I.S. Ability, 01-01-1955 (64 years of age)**

**Assessment date: 13-02-2019**

**Status:** pending **Registration date**: 15-02-2019 **Medical expert**: Dr. Zhivago **Vocational expert**: Mr. Match **FML version**: 4, November 2002

| Client information | Assessment information | **Pre-selection** | Final selection | Results of final selection | Questions/answers | Printout/close case | Further notes |
|---|---|---|---|---|---|---|---|

**Selected**

| Degree of disability category | # Standard Occupation classification codes | # job types | # jobs in database |
|---|---|---|---|
| < 35 | 119 | 2028 | 7700 |
| 34-45 | 13 | 60 | 487 |
| 45-55 | 1 | 17 | 96 |
| 55-65 | 3 | 40 | 1030 |
| 80-100 | 1 | 3 | 3 |

**Rejected**

| Criterion | Selection criterion | # job types | # jobs in database |
|---|---|---|---|
| Too few jobs in database | < 3 | 39 | 43 |
| Education level | 4 | 1778 | 7257 |
| Medical hours restriction per week | Hours: 30 | 945 | 6193 |
| Work pattern – periods per day | Night: no | 917 | 7082 |

| **Return to main page** | **Confirm selection** |
|---|---|

FIGURE A1. FICTIONAL EXAMPLE OF PRE-SELECTION JOB MATCHES FOUND BY ALGORITHM FROM EMPLOYMENT AGENCY'S DATABASE

The occupational assessor must select at least three job types that are feasible for the applicant. Selection must be made in order from higher to lower average earnings. The assessor can decide to skip particular job types despite being functional matches. Each of these job types must have at least three job positions. If more than three job types are feasible, then some are selected as "spares" to be used in case of appeal. With an appeal, if one of the first three selected job types is ruled by a judge to be infeasible, then it can be replaced by a spare, which would result in a higher degree of disability and benefit entitlement. Spares may also be used if one of the first three listed job types are considered by the assessor, after consultation with the physician, to be infeasible for the applicant. This would lead to re-calculation of the degree of disability and benefit entitlement. This can continue interactively, giving the occupational assessor the opportunity to observe the impact of different job rejections on benefit entitlement. After ending the process, the assessor must explain in a note why any of the job matches found by the algorithm were rejected.

Figure A2 gives a fictional example of the calculation of degree of disability using the jobs selected by the assessor from the job matches found by the algorithm.

**Client dossier: Mr. D.I.S. Ability, 01-01-1955 (64 years of age)**

**Assessment date: 13-02-2019**

**Status:** pending **Registration date**: 15-02-2019 **Medical expert**: Dr. Zhivago **Vocational expert**: Mr. Match **FML version**: 4, November 2002

| Client information | Assessment information | Pre-selection | Final selection | **Results of final selection** | Questions/answers | Printout / close case | Further notes |
|---|---|---|---|---|---|---|---|

**Overview final selection of Standard Occupation Classification (SBC) codes and respective job types**

| SBC code | Job-id | Job name | Hours p/w | Reduction factor | Hourly wage | # jobs in database | Status | Q & A |
|---|---|---|---|---|---|---|---|---|
| 111160 | 2271.0008.001 | Dressmaker | 33,0 | 1,0000 | € 12,59 | 1 | Selected | |
| 111160 | 3441.0381.035 | Worker in clothes workspace | 32,0 | 1,0000 | € 12,17 | 1 | Selected | |
| 111160 | 9832.0005.018 | Cloth repairing | 32,0 | 1,0000 | € 11,70 | 1 | Selected | |
| 111160 | 2271.0007.003 | Curtain maker | 32,0 | 1,0000 | € 11,10 | 3 | Selected | |
| 516130 | 7611.0000.010 | Sales employee | **20,0** | **0,6250** | **€ 9,95** | 3 | Selected | |
| 111230 | 9548.9999.004 | Driver / deliverer | 32,0 | 1,0000 | € 9,72 | 3 | Selected | |
| 263020 | 9011.0007.011 | Maintenance | 32,0 | 1,0000 | € 15,68 | 10 | Excluded | |
| 372051 | 9311.0314.014 | Care assistant | 32,0 | 1,0000 | € 12,19 | 18 | Excluded | |
| 468080 | 2632.9997.008 | Offset printing worker | 35,0 | 1,0000 | € 16,45 | 6 | Excluded | |

**Calculation final outcomes**

**By SBC code**

| **SBC code** | **Hourly wage** | **Hours p/w** |
|---|---|---|
| 111160 | € 11,935 | 33,0 |
| 516130 | € 9,950 | **20,0** |
| 111230 | € 9,720 | 32,0 |

**Hourly wages**

| | |
|---|---|
| Pre-disability wage | € 15,00 |
| Median assessed wage | € 9,950 |

**Working hours**

| | |
|---|---|
| Pre-disability job: | 32,00 |
| Minimum: | 20,00 |
| Reduction factor: | 0,6250 |

| | |
|---|---|
| Earning capacity per hour | € 6,22 |
| Earning capacity per month | € 865,65 |
| Degree of disability | 58,54 % |
| Degree of disability category | 55-65 |

| **Return to main page** | **Confirm selection** |
|---|---|

FIGURE A2. FICTIONAL EXAMPLE OF DEGREE OF DISABILITY CALCULATION FROM POST-SELECTION JOB MATCHES

There are three ways in which the occupational assessor could potentially influence final benefit entitlement. First, if the assessor perceives the functional limitations score to be incomplete or an inaccurate reflection of an applicant's functional capacity, the physician can be asked to reconsider. For a small fraction of applications, the physician makes changes that facilitate matching of functional capacity to additional jobs. This generally leads to an increase in the number of feasible jobs and reduces the degree of disability. Second, the occupational assessor has discretion in choosing a final set of feasible job types. This process, which may lead to a higher degree of disability as a result of both the number and types of jobs chosen, is relatively straightforward as job types are listed by the benefit interval each would produce (Figure A2). Third, the occupational assessor can reconsider an applicant's capacity for a particular job. The assessor could be prompted to do this by doubt about whether the job selection would be upheld in an appeal court.

## References


**Koning, P. and J. Van Sonsbeek.** 2017. "Making Disability Work? The Effects of Financial Incentives on Partially Disabled Workers." *Labour Economics*, 47, 202-15.

## 2. Appendix B. Data requests and access

All administrative records used in this paper were requested and obtained from the Dutch Employee Insurance Agency (UVW) via a formal data request to their Knowledge Department. There is no public access to the data without a formal request to UWV. All analyses of the administrative records were conducted via secure servers at the UVW.

**Disability Insurance Claims data**

We had access to all applications for DI benefits from 2003 to 2017. Each record contains information on the applicant's demographics, diagnosis and functional limitations score from the medical examination, and the disability of degree (DI entitlement) outcome of the application. There is also information on the insurance agency office that handled the application.

**Functions data**

For 2003 to 2017, we had access to another dataset containing a record for each application of the jobs, number and types of positions, and respective wages and hours extracted from the agency's database by the occupational assessor using the algorithm search for job matches. This information was used to calculate the final degree of disability that determined the benefit entitlement of the applicant.

**CBBS data**

For applications in the period 2011 to 2018, we had access to a dataset that provided information on the ID of the examining physician and the occupational assessor as well as the potential outcome in each assessment.

## 3. Appendix C. Additional Figures and Tables

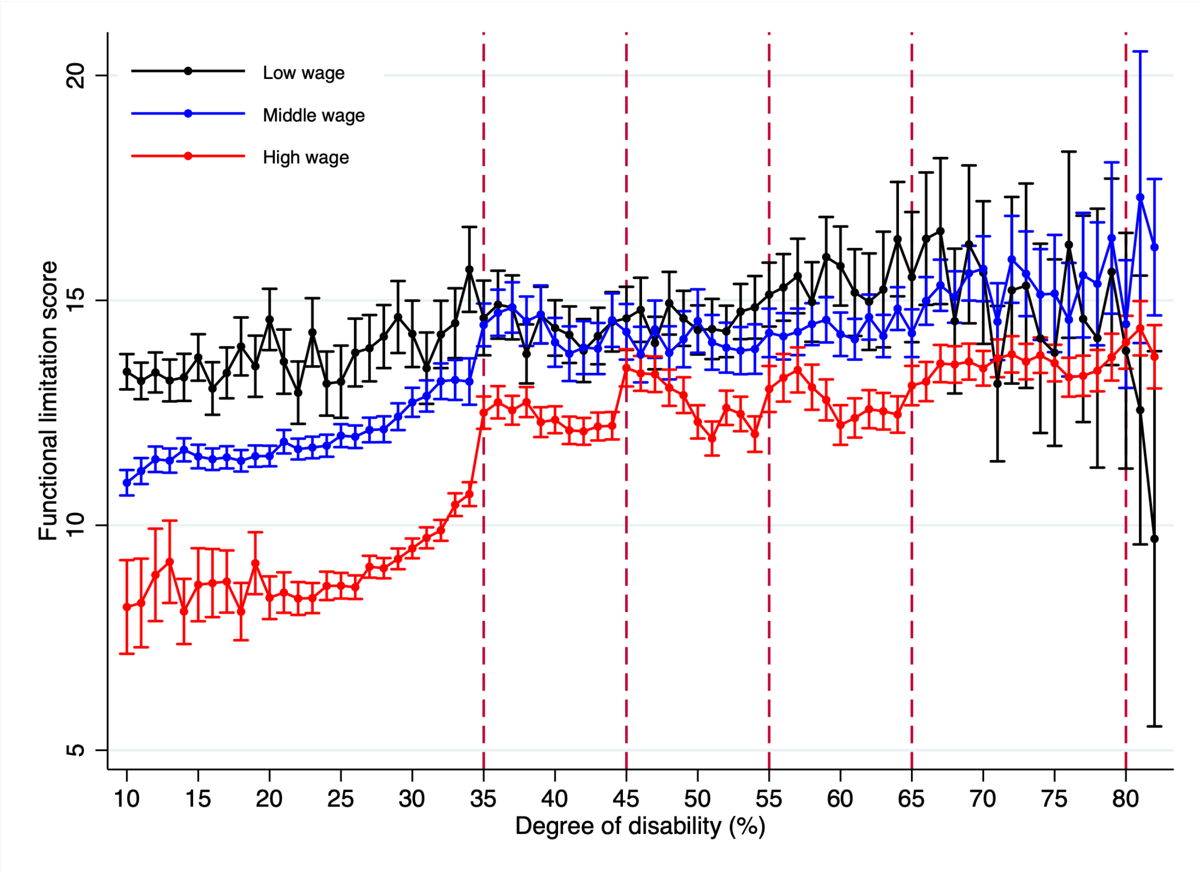


FIGURE C1. MEAN FUNCTIONAL LIMITATIONS SCORE BY DEGREE OF DISABILITY STRATIFIED BY WAGE TERCILE GROUP

*Notes*: A square indicates an OLS estimate of mean functional limitations score at the respective percentage point of the degree of disability. Interval lines are 95% confidence intervals. Stratification by tercile groups of pre-disability wage. N=192,375 applications. Dashed lines show DI entitlement thresholds.

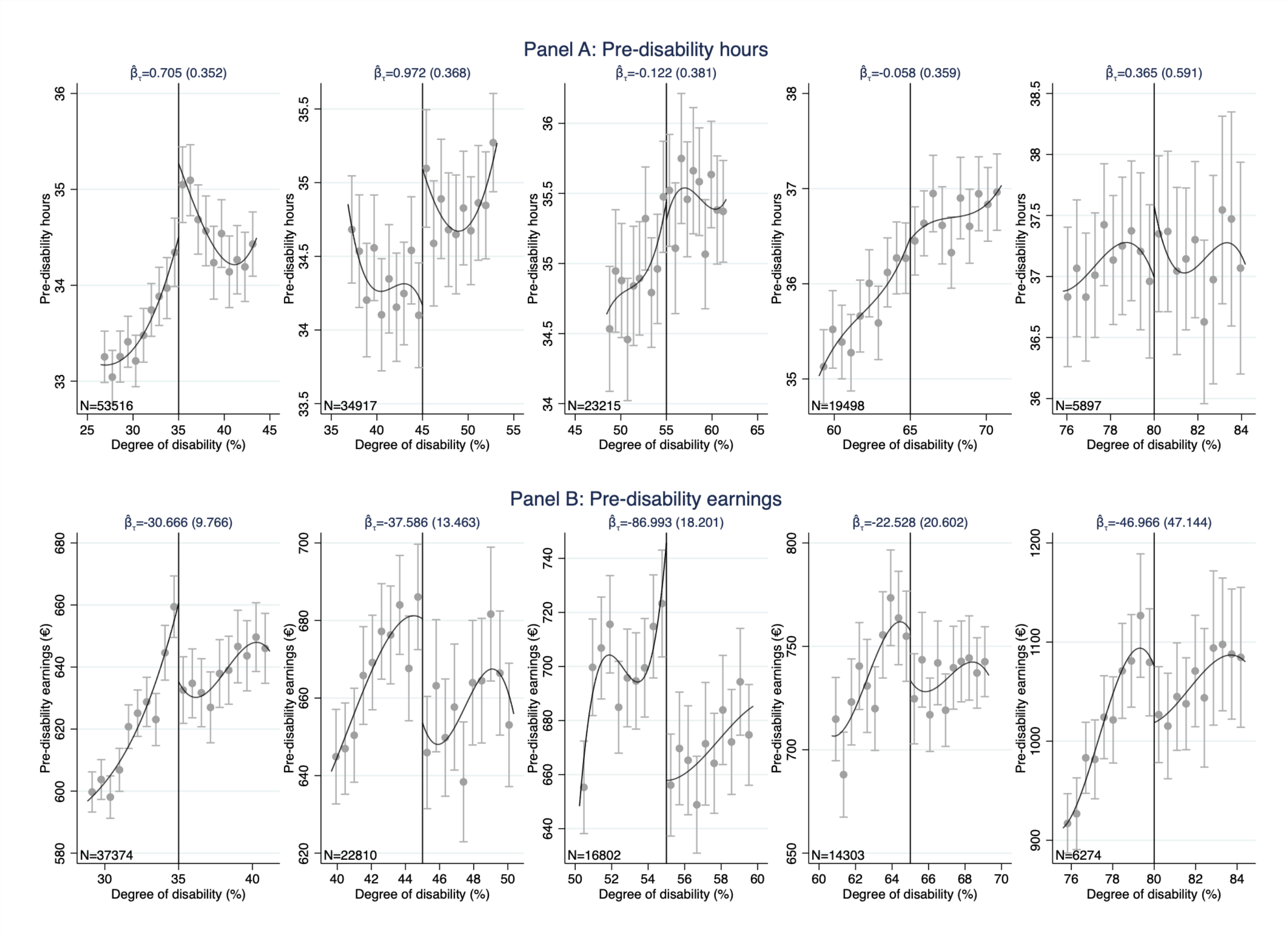


FIGURE C2. DISCONTINUITIES IN PRE-DISABILITY HOURS AND EARNINGS AT DISABILITY INSURANCE ENTITLEMENT THRESHOLDS

*Notes*: Bias-corrected robust regression discontinuity with a different third order polynomial on each side of the respective threshold and an optimal threshold-specific bandwidth. Dots represent bin averages (10 bins on each side of threshold). Interval lines show 95% confidence intervals.

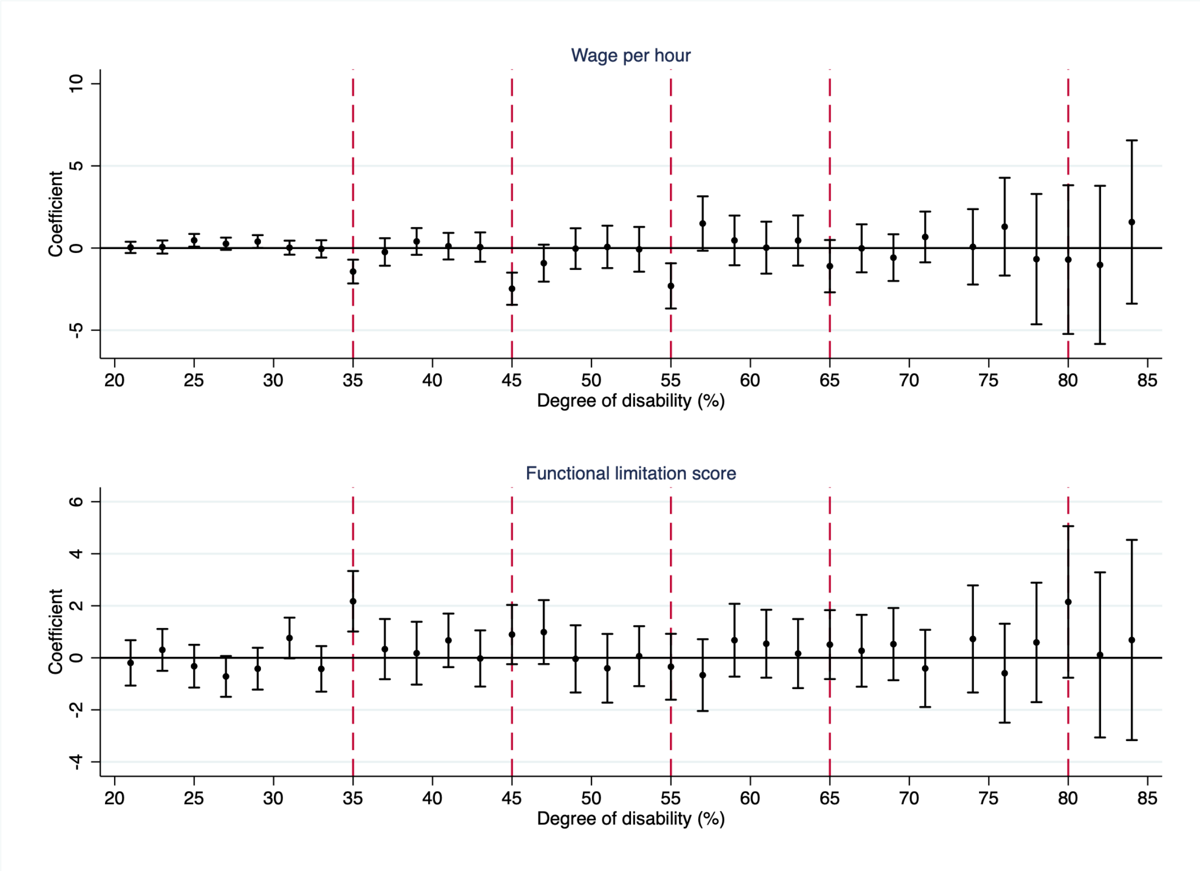


FIGURE C3. DISCONTINUITIES IN PRE-DISABILITY WAGE AND FUNCTIONAL LIMITATIONS SCORE AT DIFFERENT DEGREES OF DISABILITY

*Notes*: Each dot is a point estimate of $\beta_\tau$ from a regression specified as (2) in the paper. A separate bias-corrected robust regression discontinuity is estimated for each value of $T_\tau$ in {21%, 23%, …, 71%, 74%, 76%, …, 84%}. Each regression allows a different third order polynomial on each side of $T_\tau$. The common bandwidth is set to 2% for each regression. Interval lines show 95% confidence intervals. Dashed lines indicate thresholds at which DI entitlement changes.

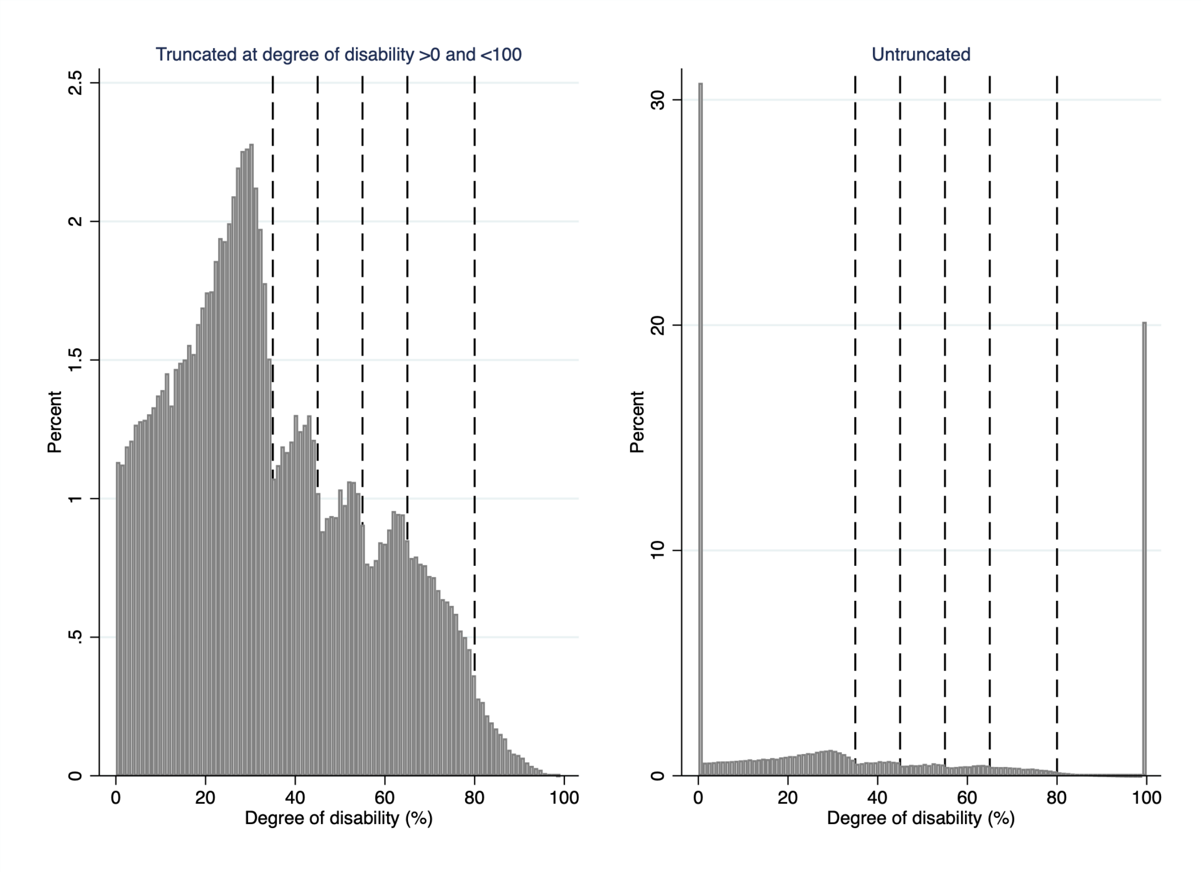


FIGURE C4. DISABILITY INSURANCE APPLICATIONS BY DEGREE OF DISABILITY

*Notes*: In Panel A, to focus on thresholds critical to benefit entitlement, we truncate at degree of disability above 0 (excluding ~30%) and below 100 (excluding ~20%). N=192,538 applications in this panel. Panel B includes all applications (N=387,917). See Table C3 for bunching estimates.

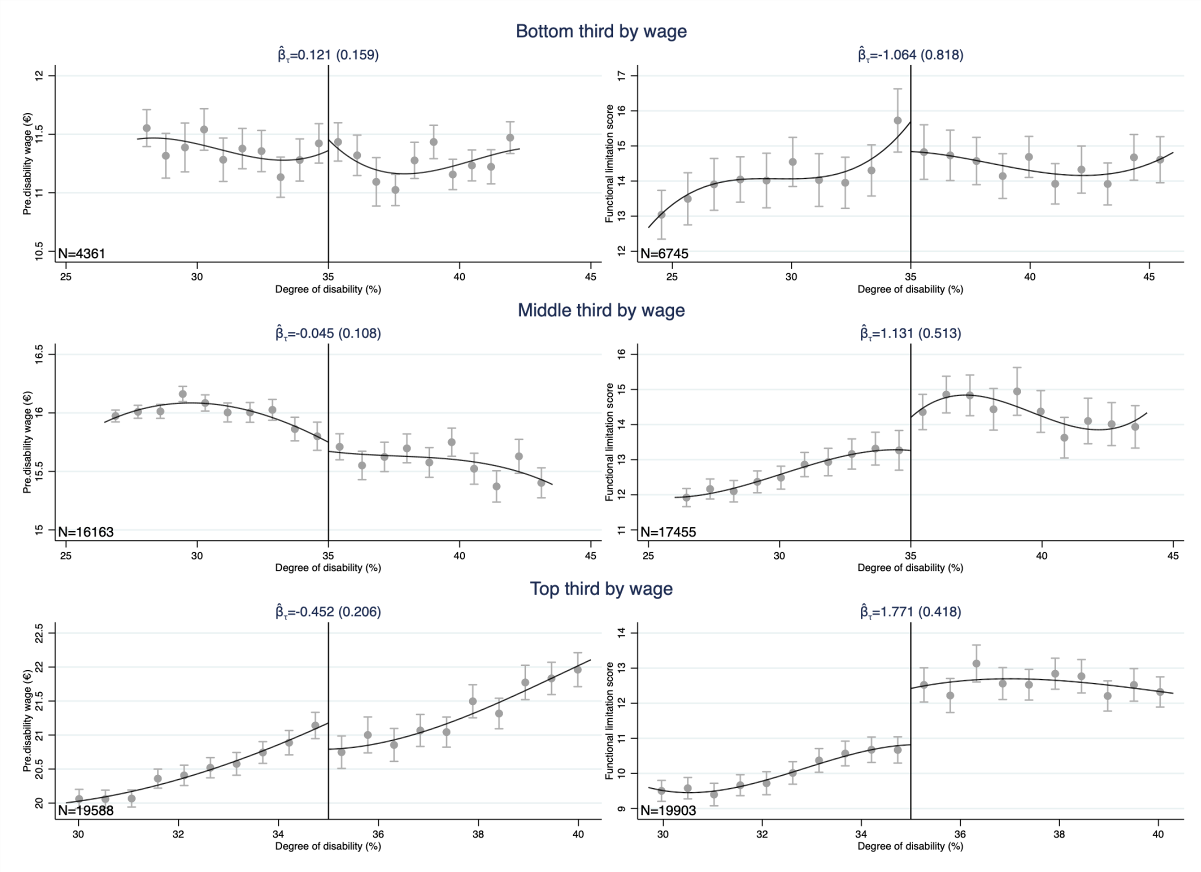


FIGURE C5. PRE-DISABILITY WAGE AND FUNCTIONAL LIMITATIONS SCORE BY DEGREE OF DISABILITY AT THRESHOLD TO QUALIFY FOR LOWEST DI BENEFIT STRATIFIED BY WAGE TERCILE GROUP

Notes: Dependent variable in panels A, C, and E is the pre-disability hourly wage. Dependent variable in panels B, D, and F is functional limitations score. Samples for panels A & B, C & D, and E & F include applicants in the bottom, middle, top third of the pre-disability wage distribution, respectively. Each panel shows estimates from a bias-corrected robust regression discontinuity as in equation (1) in paper, with a different third order polynomial on each side of the threshold and a common optimal bandwidth. Dots represent bin averages (10 bins on each side of threshold). Interval lines show 95% confidence intervals.

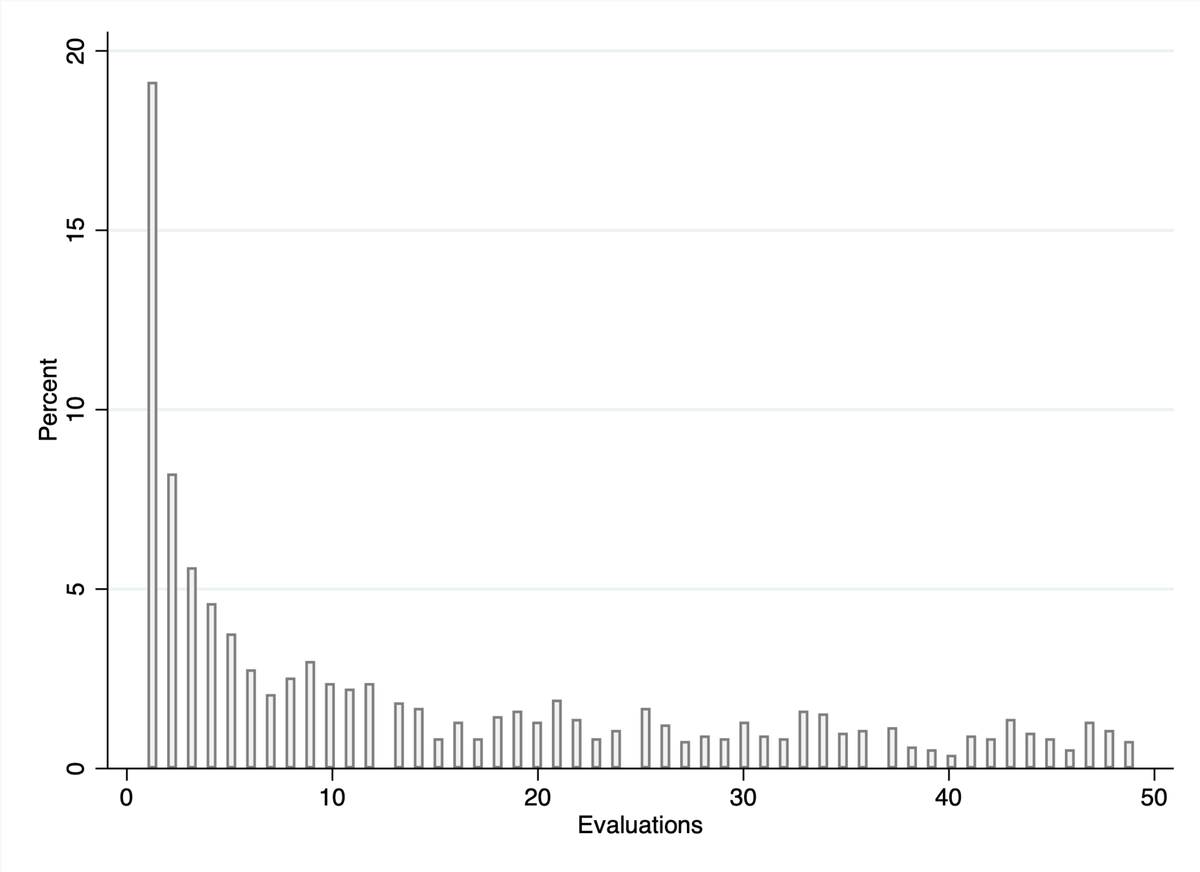


FIGURE C6. DISTRIBUTION OF NUMBER OF APPLICATIONS HANDLED BY ASSESSORS FOR THOSE HANDLING LESS THAN 50 IN TOTAL

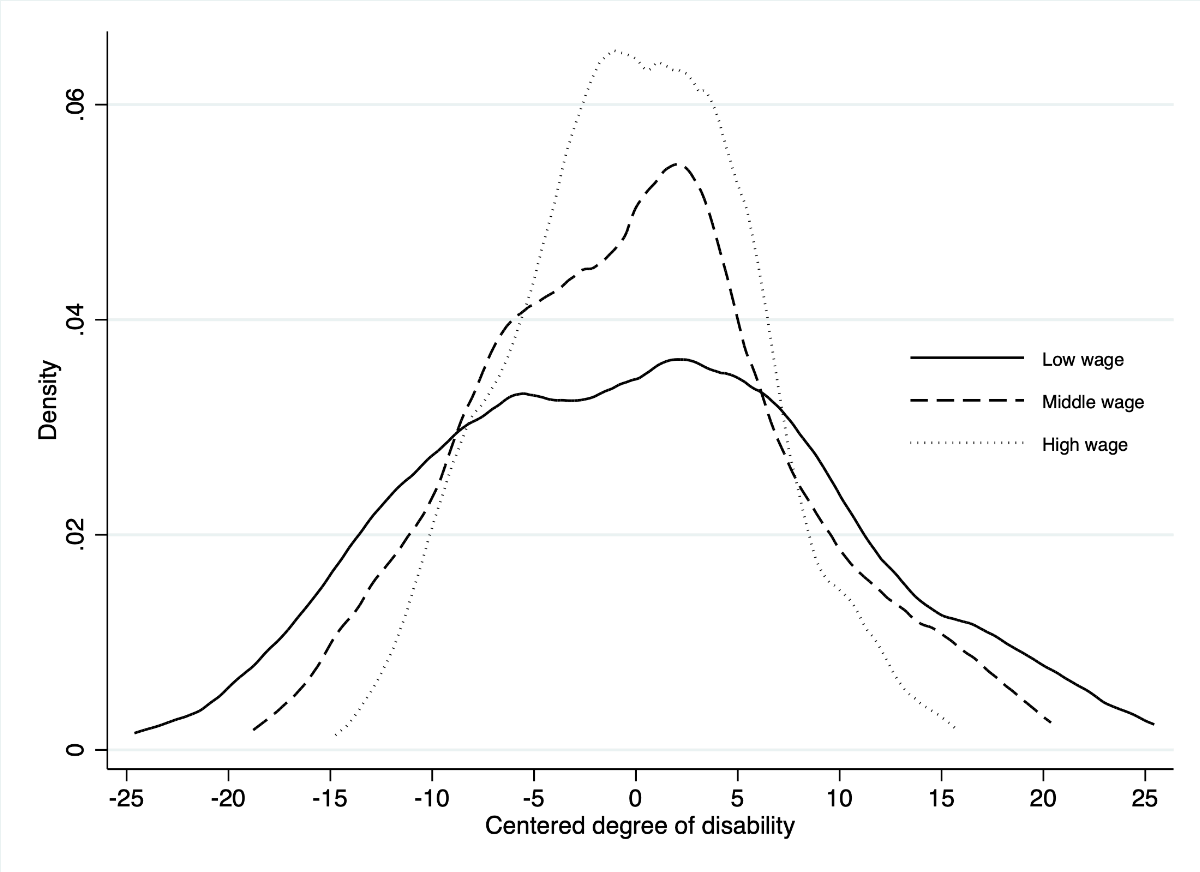


FIGURE C7. DISTRIBUTIONS OF (MEAN-CENTERED) ASSESSOR FIXED EFFECTS ON DEGREE OF DISABILITY BY PRE-DISABILITY WAGE TERCILE GROUP

*Note*. Distributions of $\hat{\lambda}_j^g - 1/n_J \sum_j \hat{\lambda}_j^g$, where $\hat{\lambda}_j^g$ is the OLS estimate of $\lambda_j^g$ from (3) and $n_J$ is the of assessors.

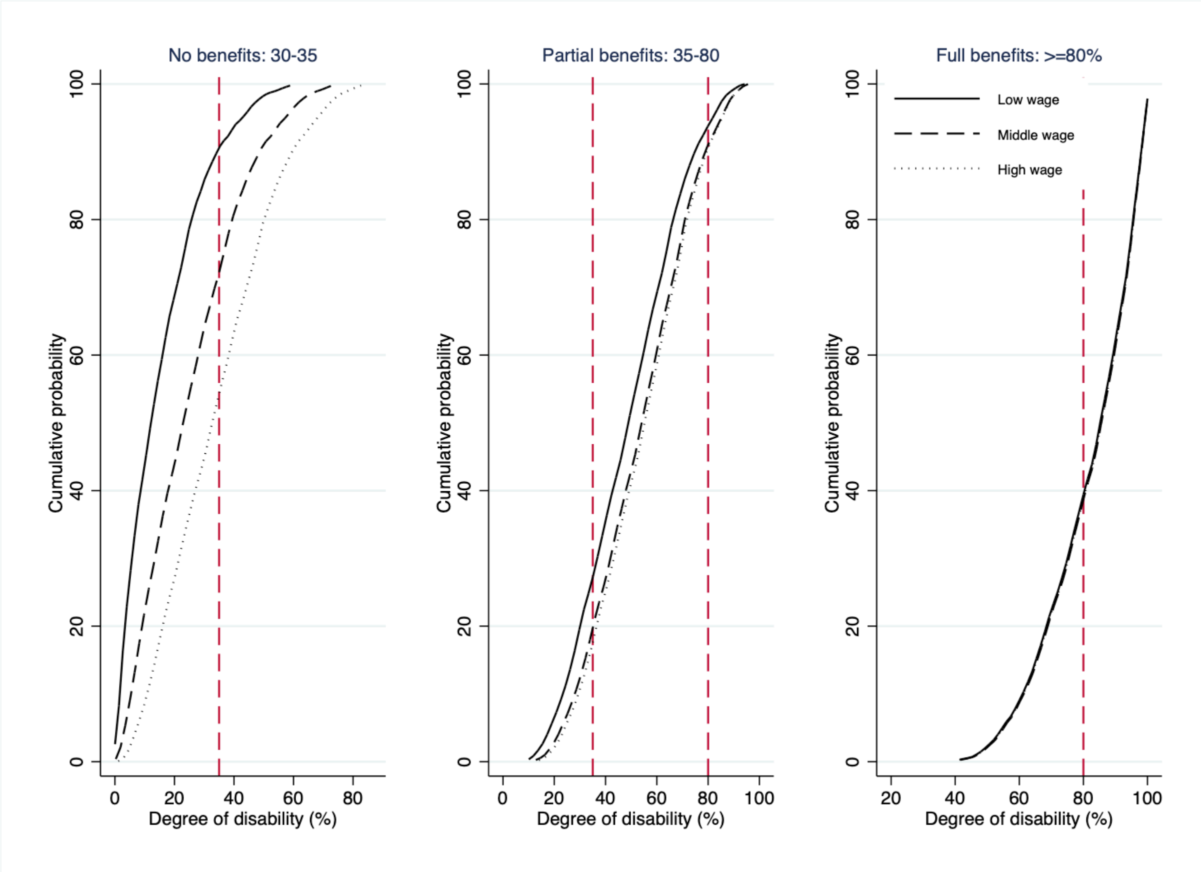


FIGURE C8. CUMULATIVE DEGREE OF DEGREE-OF-DISABILITY DISTRIBUTIONS GENERATED BY ASSESSOR FIXED EFFECTS, STRATIFIED BY ACTUAL DI AWARD AND PRE-DISABILITY WAGE TERCILE GROUP

Note. For each actual DI award category and wage group, the figure plots the average value of $Pr\left(\widetilde{DD}_{ikrt}^{g} < X\right)$ at $X \in \{1, \ldots, 100\}$, where $\widetilde{DD}_{ikrt}^{g}$ is obtained for each applicant from (4). Actual DI award categories no benefit, partial benefit, and full benefit, correspond to actual DD < 35%, 35% ≤ DD < 80%, and DD ≥80%, respectively. Low-wage, middle-wage, and high-wage are the bottom, middle, and top third of the pre-disability wage distribution, respectively. Dashed lines show the DI entitlement thresholds.

TABLE C1. CHARACTERISTICS OF DISABILITY INSURANCE APPLICATIONS, 2006-2017

| | Mean (SD) |
|---|---|
| Male | 0.48 |
| Age | 46.02 (10.94) |
| | |
| Education | |
| Basic education | 0.26 |
| Vocational high school | 0.48 |
| Professional high school | 0.15 |
| College | 0.11 |
| | |
| Employment | |
| Permanent contract | 0.48 |
| Temporary contract/unemployed | 0.52 |
| Hourly wage (€) | 16.54 (6.59) |
| Hours per week | 32.30 (9.23) |
| | |
| Main diagnosis | |
| Musculoskeletal | 0.35 |
| Psychiatric | 0.30 |
| Cardiovascular | 0.07 |
| Neurological | 0.07 |
| Urinary | 0.04 |
| Respiratory | 0.03 |
| Digestive | 0.03 |
| General | 0.02 |
| Others | 0.07 |
| | |
| Functional limitation score | 12.92 (8.53) |
| | |
| Degree of disability (%) | 37.45 (38.01) |
| 0 to 34 | 0.59 |
| 35 to 44 | 0.06 |
| 45 to 54 | 0.05 |
| 55 to 64 | 0.04 |
| 65 to 79 | 0.05 |
| 80 to 100 | 0.21 |
| | |
| Observations | 387899 |

*Notes*: Excludes applications categorized as fully disabled on basis of medical examination or functional limitations that precluded matching with any jobs without occupation assessment. Values are at application for DI.

## TABLE C2. DISCONTINUITIES IN DETERMINANTS OF DISABILITY INSURANCE ENTITLEMENT AT THRESHOLDS - ROBUSTNESS

| | | Main estimate | Bandwidth | | Function | |
|---|---|---|---|---|---|---|
| | | | 1 pp | 4 pp | Linear | Quadratic |
| | | (1) | (2) | (3) | (4) | (5) |
| *Panel A: 35% threshold* | | | | | | |
| Wage per hour (€) | Coeff | -1.365 | -1.467 | -1.171 | -1.322 | -1.185 |
| | *SE* | *0.226* | *0.328* | *0.166* | *0.119* | *0.172* |
| Hours per week | Coeff | 0.705 | 0.337 | 0.913 | 0.876 | 0.821 |
| | *SE* | *0.352* | *0.503* | *0.272* | *0.187* | *0.268* |
| Earnings per week (€) | Coeff | -30.666 | -40.899 | -19.353 | -26.598 | -28.232 |
| | *SE* | *9.766* | *14.067* | *7.169* | *5.098* | *7.396* |
| Functional limitations score | Coeff | 1.655 | 1.893 | 1.470 | 1.801 | 1.551 |
| | *SE* | *0.306* | *0.446* | *0.226* | *0.160* | *0.231* |
| | N | 40152 | 18060 | 80056 | 40152 | 40152 |
| *Panel B: 45% threshold* | | | | | | |
| Wage per hour (€) | Coeff | -1.637 | -2.257 | -1.246 | -1.686 | -1.601 |
| | *SE* | *0.296* | *0.413* | *0.220* | *0.162* | *0.230* |
| Hours per week | Coeff | 0.972 | 1.224 | 0.806 | 0.767 | 0.856 |
| | *SE* | *0.368* | *0.517* | *0.283* | *0.201* | *0.284* |
| Earnings per week (€) | Coeff | -37.586 | -64.938 | -27.327 | -41.025 | -33.330 |
| | *SE* | *13.463* | *18.826* | *10.000* | *7.347* | *10.404* |
| Functional limitations score | Coeff | 0.984 | 0.998 | 0.650 | 1.029 | 0.887 |
| | *SE* | *0.387* | *0.537* | *0.282* | *0.208* | *0.298* |
| | N | 20081 | 9891 | 40225 | 20081 | 20081 |
| *Panel C: 55% threshold* | | | | | | |
| Wage per hour (€) | Coeff | -2.197 | -2.321 | -1.968 | -1.870 | -2.141 |
| | *SE* | *0.499* | *0.696* | *0.375* | *0.273* | *0.383* |
| Hours per week | Coeff | -0.122 | -0.443 | 0.000 | 0.199 | -0.003 |
| | *SE* | *0.381* | *0.530* | *0.291* | *0.213* | *0.296* |
| Earnings per week (€) | Coeff | -86.993 | -93.877 | -67.663 | -64.110 | -70.404 |
| | *SE* | *18.201* | *25.452* | *14.000* | *10.095* | *14.081* |
| Functional limitations score | Coeff | 0.607 | -0.129 | 0.380 | 0.738 | 0.597 |
| | *SE* | *0.418* | *0.585* | *0.305* | *0.225* | *0.320* |
| | N | 17529 | 8634 | 35460 | 17529 | 17529 |
| *Panel D: 65% threshold* | | | | | | |
| Wage per hour (€) | Coeff | -0.644 | -1.011 | -1.124 | -1.339 | -1.056 |
| | *SE* | *0.544* | *0.761* | *0.427* | *0.299* | *0.421* |
| Hours per week | Coeff | -0.058 | 0.128 | -0.123 | 0.144 | 0.062 |
| | *SE* | *0.359* | *0.507* | *0.279* | *0.196* | *0.275* |
| Earnings per week (€) | Coeff | -22.528 | -25.787 | -31.624 | -41.438 | -32.558 |
| | *SE* | *20.602* | *28.564* | *16.325* | *11.514* | *16.088* |
| Functional limitations score | Coeff | 0.133 | -0.100 | -0.034 | 0.331 | 0.057 |
| | *SE* | *0.383* | *0.533* | *0.288* | *0.208* | *0.295* |
| | N | 20736 | 10752 | 39383 | 20736 | 20736 |
| *Panel E: 80% threshold* | | | | | | |
| Wage per hour (€) | Coeff | -1.820 | -0.693 | -2.842 | -2.925 | -2.548 |
| | *SE* | *1.195* | *1.637* | *0.979* | *0.695* | *0.954* |
| Hours per week | Coeff | 0.365 | 0.259 | 0.282 | 0.039 | 0.419 |
| | *SE* | *0.591* | *0.799* | *0.479* | *0.341* | *0.464* |
| Earnings per week (€) | Coeff | -46.966 | -22.731 | -90.691 | -106.034 | -77.476 |
| | *SE* | *47.144* | *66.025* | *38.827* | *27.997* | *37.901* |
| Functional limitations score | Coeff | 0.697 | 2.277 | 0.381 | 0.426 | 0.321 |
| | *SE* | *0.751* | *1.071* | *0.590* | *0.415* | *0.577* |
| | N | 6426 | 3184 | 13170 | 6426 | 6426 |

*Notes*: Bias-corrected robust regression discontinuity estimates of $\beta_\tau$ from model (1) in paper. Column (1) estimates are those given in Figures 2 for the wage and functional limitations score. They are obtained with each of $f_{\tau L}(DD_i)$ and $f_{\tau R}(DD_i)$ specified as a different third-order polynomial, the same (coverage error probability) optimal bandwidth on each side of the threshold, and the dependent variable (Y) in levels. Columns (2) and (3) deviate from this specification by using a bandwidth of 1 percentage point (pp) and 4 pp, respectively. Columns (4) and (5) deviate from the column (1) specification by specifying each of $f_{\tau L}(DD_i)$ and $f_{\tau R}(DD_i)$ as linear and quadratic, respectively. SE = standard error. Wage, hours, and earnings are prior to application for DI.

TABLE C3. BUNCHING ESTIMATES OF MISSING MASS OF APPLICATIONS AT DISABILITY INSURANCE ENTITLEMENT THRESHOLDS

| | Degree of disability threshold for DI entitlement | | | | |
|---|---|---|---|---|---|
| | 35% | 45% | 55% | 65% | 80% |
| A. *Overall* | -0.0013 | -0.0015 | -0.0010 | -0.0003 | -0.0002 |
| | (0.0003) | (0.0003) | (0.0003) | (0.0004) | (0.0004) |
| N | 28860 | 17944 | 10809 | 7211 | 1314 |
| B. *Wage tercile groups* | | | | | |
| Low | -0.0002 | -0.0002 | 0.0010 | 0.0003 | 0.0001 |
| | 0.0004 | 0.0003 | 0.0005 | 0.0003 | 0.0002 |
| N | 2644 | 7540 | 2438 | 385 | 61 |
| Middle | 0.0018 | 0.0008 | -0.0006 | -0.0003 | 0.0000 |
| | 0.0007 | 0.0005 | 0.0007 | 0.0007 | 0.0003 |
| N | 4677 | 3766 | 3301 | 4162 | 222 |
| High | -0.0064 | -0.0057 | -0.0034 | -0.0005 | 0.0004 |
| | 0.0010 | 0.0009 | 0.0008 | 0.0009 | 0.0011 |
| N | 18610 | 10145 | 6022 | 5260 | 1775 |
| C. *Functional limitations tercile groups* | | | | | |
| Low | -0.0026 | -0.0013 | -0.0007 | -0.0006 | -0.0000 |
| | 0.0006 | 0.0004 | 0.0004 | 0.0003 | 0.0004 |
| N | 9030 | 4615 | 4312 | 4757 | 722 |
| Middle | -0.0014 | -0.0021 | -0.0011 | -0.0003 | 0.0000 |
| | 0.0005 | 0.0005 | 0.0005 | 0.0005 | 0.0005 |
| N | 18782 | 12191 | 7150 | 5276 | 900 |
| High | -0.0001 | -0.0006 | -0.0001 | -0.0004 | 0.0006 |
| | 0.0004 | 0.0003 | 0.0006 | 0.0007 | 0.0007 |
| N | 12169 | 11153 | 2697 | 1964 | 420 |

*Notes*: Table gives estimates of the proportionate shortfall from the expected number of applications at each degree of disability threshold at which DI entitlement changes. Panel A gives estimates without any stratification. Panels B and C give estimates are from samples stratified by wage tercile and functional limitations score terciles, respectively. Estimates are obtained using the non-parametric density discontinuity test of Cattaneo et al. (2021). Estimates are obtained using optimal bandwidths on each side of the threshold. Robust standard errors in parentheses.

TABLE C4. DISABILITY INSURANCE APPLICATIONS HANDLED BY ASSESSORS

| | Mean | SD |
|---|---|---|
| Overall | 119.53 | 149.97 |
| Wage tercile group of applicants: | | |
| Low | 39.99 | 50.12 |
| Middle | 40.36 | 50.68 |
| High | 39.18 | 51.14 |
| Number of assessors | | |
| Total | 2635 | |
| Handled ≥ 10 applications per wage group and ≥50 in total | 1327 | |

TABLE C5. ASSESSOR FIXED EFFECTS ON DEGREE OF DISABILITY BY PRE-DISABILITY WAGE TERCILE GROUP

| | Mean | SD | Median | IQR |
|---|---|---|---|---|
| Wage tercile group of applicants | | | | |
| Low | -25.27 | 10.17 | -25.34 | 14.71 |
| Middle | -10.52 | 7.75 | -10.42 | 10.34 |
| High | 8.26 | 5.70 | 8.31 | 7.96 |

*Notes*. Summary statistics of distributions of least squares estimates of $\lambda_j^g$ from regression models (3) in paper. A separate regression is estimated using applications in each wage tercile group. SD=standard deviation. IQR=inter-quartile range.

## 4. Appendix D. Exercise of discretion in assessing DI applications

This appendix presents evidence that the discontinuities in pre-disability wages and in the functional limitations score observed at entitlement thresholds do not result from the medical re-examination. Furthermore, we explore how occupational assessors use discretion in selection and rejection of job matches to favor, perhaps subconsciously, certain types of applicants. Finally, this appendix shows the difference in assessors' use of discretion close to the thresholds for low-waged and high-waged applicants.

There is greater scope for the exercise of discretion in the occupational assessment of a DI application than there is in the medical examination. Only the occupational assessor observes how decisions made during an evaluation would affect an applicant's degree of disability and, consequently, DI entitlement. The assessor may ask for another medical examination, but this is relatively rare (Table D1). Discontinuities in pre-disability wages and the functional limitations score at the 35% degree of disability threshold at which applicants qualify for any DI benefit are larger, and only significant, for applications with just one medical examination (Figure D1). This difference suggests that the discontinuities do not arise because of medical re-examinations.

TABLE D1. DI APPLICATIONS BY NUMBER OF EVALUATIONS, 2011 TO MID-2017

| | Number | Proportion |
|---|---|---|
| Number of evaluations | | |
| 1 | 235253 | 0.83 |
| 2 | 39365 | 0.14 |
| 3 | 7520 | 0.03 |
| 4+ | 2116 | 0.01 |
| Number of medical examinations | | |
| 1 | 246316 | 0.87 |
| 2 | 31883 | 0.11 |
| 3 | 4880 | 0.02 |
| 4+ | 1175 | 0.00 |
| Number of occupational assessments | | |
| 0 | 67123 | 0.24 |
| 1 | 189564 | 0.67 |
| 2 | 23564 | 0.08 |
| 3 | 3303 | 0.01 |
| 4+ | 700 | 0.00 |
| N | 285305 | |

*Notes*. The unit of observation is an application for disability insurance. The table covers all applications from 2011 to mid-2017. An evaluation is a decision on DI entitlement. An application can be evaluated more than once if there is an appeal by the applicant to the evaluation result. The medical examinations are conducted by an insurance agency physician. The occupation assessor can however ask the physician to conduct another examination. Applications with no occupational assessment are those for which full disability can be established from the medical examination or diagnosis without the need to calculate degree of disability based on job matches. More than one occupational assessment can be conducted because the occupational assessor receives a new medical examination.

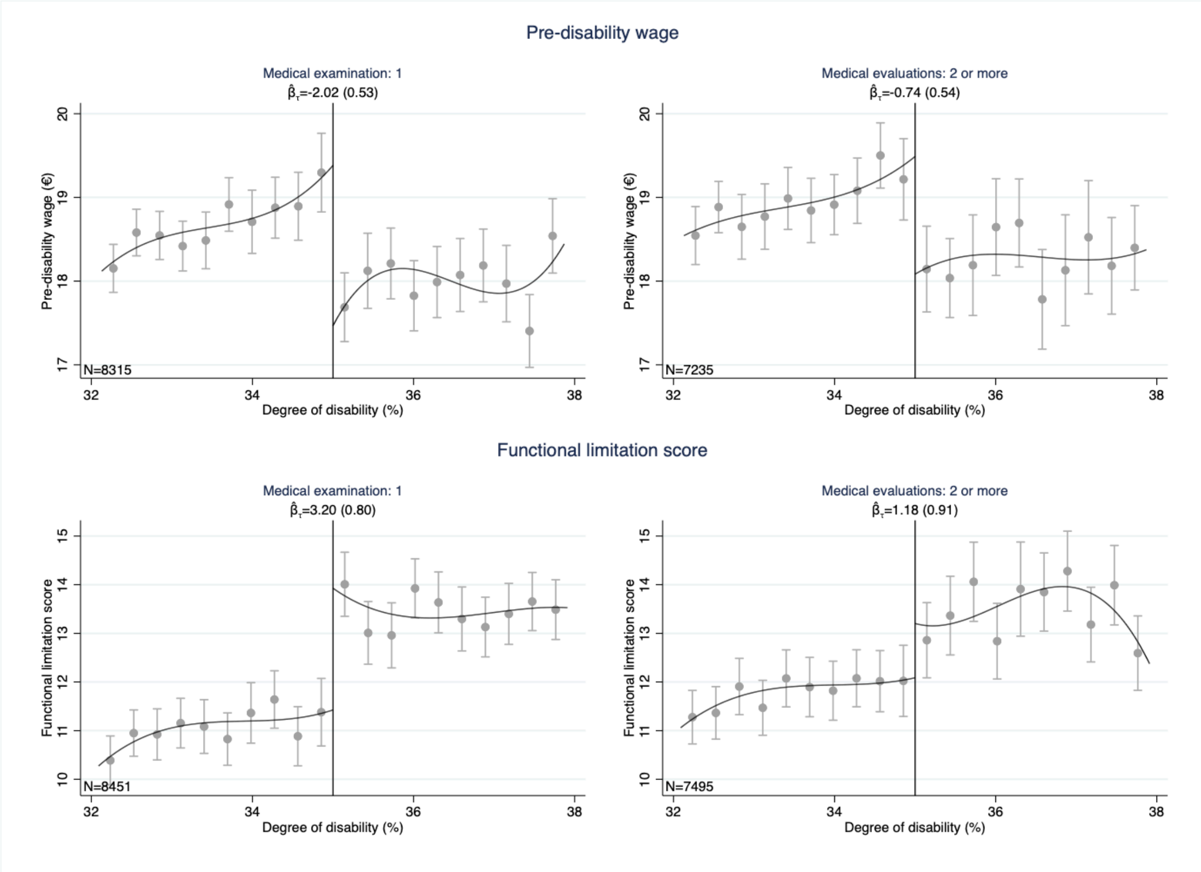


FIGURE D1. PRE-DISABILITY WAGE AND FUNCTIONAL LIMITATIONS SCORE BY DEGREE OF DISABILITY AT THRESHOLD FOR ANY DI BENEFIT STRATIFIED BY NUMBER OF MEDICAL EXAMINATIONS

*Notes*: Bias-corrected robust regression discontinuity (model (1) in paper) with a different third order polynomial on each side of the respective threshold and using the same optimal threshold-specific bandwidth for each outcome. Dots represent bin averages. Interval lines show 95% confidence intervals. $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

For each applicant, the occupational assessor selects job types from all the job matches that the algorithm finds in the insurance agency's database. According to guidelines, the assessor should select jobs in increasing order of degree of disability. After selecting job matches, the assessor can then rule out some (higher paying) jobs because they are considered infeasible for the applicant. Both through the selection of job matches and by ruling some jobs to be infeasible, the assessor can exercise discretion in the calculation of the degree of disability.

Our ability to identify the exercise of discretion is limited by the fact that we do not observe all the job matches listed by the algorithm. However, for all applications from 2011 to mid-2017, we do observe all the matches in the selection made by the assessor. In this appendix, we use these data to examine the exercise of discretion and how it differs with applicant characteristics.

We restrict attention to applicants with a final degree of disability in the region of the 35% threshold that gives entitlement to any benefit. For each of these applicants, we calculate the proportion of job matches in the selection made by the assessor that would result in a degree of disability greater than or equal to 35%. Keep in mind that the default is to select the matches that result in the lowest degree of disability. We use the proportion of selected job matches that give a degree of disability of at least 35% as the dependent variable in a regression discontinuity analysis of the form given by equation (2) in the paper. The proportion of selected matches that give DD ≥ 35% will be higher for applicants who are above this threshold. But the proportion need not jump discontinuously at the threshold. The degree by which it does may indicate the extent to which job matches are purposively selected to cross the threshold. We stratify the analysis by pre-disability wage tercile groups and compare discontinuities across these groups. The estimates are given in Table D2 using the logarithmic transformation of the proportion of jobs to obtain the relative jump for each group.

At the 35% threshold, the proportion of selected job matches that would place an applicant above that threshold rises discontinuously by approximately 56% for applicants in the bottom third of the wage distribution but by only 17% for those in the top third. It appears that for higher-waged applicants assessors are stricter in adhering to the guideline of selecting job matches that give the lowest degree of disability.

TABLE D2. DISCONTINUITY IN PROPORTION OF SELECTED JOB MATCHES THAT GIVE A DEGREE OF DISABILITY ≥ 35% THRESHOLD FOR PARTIAL BENEFITS

| | (1)<br>Low-waged | (2)<br>Medium-waged | (3)<br>High-waged |
|---|---|---|---|
| $\hat{\beta}_\tau$ | 0.445* | 0.307*** | 0.163* |
| (SE) | (0.187) | (0.090) | (0.066) |
| N | 52834 | 54424 | 52775 |

Notes: Bias-corrected robust regression discontinuity estimates of $\beta_\tau$ from model (2) in paper at $T_\tau = 35\%$. Dependent variable is the log transformation of the proportion of assessor-selected job matches that would give a degree of disability ≥35%. Separate regressions estimated for applicants in the bottom, middle, and top third of the pre-disability wage distribution, with estimates given in columns (1), (2), and (3), respectively. Each regression has $f_{\tau L}(DD_i)$ and $f_{\tau R}(DD_i)$ specified as a different third-order polynomial and the same (coverage error probability) optimal bandwidth on each side of the threshold. SE = standard error.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

To investigate differential exercise of discretion in ruling jobs infeasible, we examine variation in the difference between an applicant's final degree of disability used to determine their DI benefit and the degree of disability that would have resulted from strict application of the rule to calculate it using the three selected jobs that would give the lowest degree of disability. The latter *potential degree of disability* is necessarily less than the final degree of disability used.

Table D3 gives three summary measures of the discrepancy between final and potential degree of disability in the region of each degree of disability threshold stratified by pre-disability wage tercile group. We restrict attention to applicants with a final degree of disability within two percentage points of each threshold. For example, for the 35% threshold, we take averages over all applicants with a final degree of disability from 33% to 37%.

Panel A shows the proportion of applications for whom the final and potential degrees of disability differ. That is, the proportion for whom at least one of the three highest paying jobs was ruled infeasible. This happened for 40.7% of low-waged applicants compared with 30.4% of high-waged applicants in the proximity of the 35% threshold. At each of the other thresholds, the proportion of applicants for whom higher-paying jobs were ruled infeasible is greater for low-waged applicants than it is for high-waged applicants. At all but one threshold (55%), the proportion declines monotonically in going from low- to middle- to high-waged applicants.

TABLE D3. DISCREPANCY BETWEEN FINAL DEGREE OF DISABILITY AND POTENTIAL DEGREE OF DISABILITY IF JOB MATCHES WERE NOT RULED INFEASIBLE

| | Degree of disability threshold | | | | |
|---|---|---|---|---|---|
| | 35% | 45% | 55% | 65% | 80% |
| A: Proportion with final degree of disability > potential degree of disability | | | | | |
| Low wage | 0.407 | 0.457 | 0.454 | 0.515 | 0.421 |
| | (0.015) | (0.013) | (0.014) | (0.027) | (0.051) |
| Middle wage | 0.364 | 0.368 | 0.324 | 0.394 | 0.410 |
| | (0.009) | (0.011) | (0.010) | (0.010) | (0.030) |
| High wage | 0.304 | 0.327 | 0.337 | 0.341 | 0.377 |
| | (0.006) | (0.007) | (0.008) | (0.008) | (0.010) |
| N | 10,242 | 8,176 | 6,882 | 6,612 | 2,758 |
| B: Mean (final degree of disability – potential degree of disability \| final DD > potential DD) | | | | | |
| Low wage | 7.482 | 7.141 | 4.667 | 9.030 | 8.447 |
| | (0.513) | (0.424) | (0.363) | (1.078) | (2.108) |
| Middle wage | 1.870 | 2.585 | 2.722 | 2.654 | 6.839 |
| | (0.143) | (0.227) | (0.242) | (0.221) | (1.156) |
| High wage | 0.861 | 0.981 | 1.185 | 1.559 | 1.559 |
| | (0.053) | (0.069) | (0.105) | (0.113) | (0.126) |
| N | 3,374 | 2,926 | 2,434 | 2,446 | 1,054 |
| C: Proportion of those at or above threshold who would be below if used potential degree of disability | | | | | |
| Low wage | 0.391 | 0.338 | 0.195 | 0.294 | 0.309 |
| | (0.012) | (0.011) | (0.012) | (0.024) | (0.050) |
| Middle wage | 0.095 | 0.131 | 0.139 | 0.155 | 0.318 |
| | (0.007) | (0.010) | (0.009) | (0.009) | (0.028) |
| High wage | 0.055 | 0.055 | 0.072 | 0.101 | 0.146 |
| | (0.004) | (0.005) | (0.007) | (0.007) | (0.009) |
| N | 6,223 | 4,762 | 3,937 | 3,698 | 1,754 |

Note: Final degree of disability is that used to decide the DI benefit actually paid. Potential degree of disability is the value that would have resulted from using the three highest-paying selected job matches without ruling any infeasible. Each column calculated using applicants within 2 percentage points of the respective degree of disability threshold. Panel A includes all such applicants. Panel B includes the subset with final degree of disability greater than potential degree of disability. Panel C includes the subset with final degree of disability above the respective threshold. Rows are stratified by pre-disability wage tercile groups: Low wage = bottom third of wage distribution, etc.

Panel B shows the mean difference between the final and potential degree of disability conditional on there being any difference. By construction, any difference is positive. Near the 35% threshold, by ruling higher-paying jobs infeasible, assessors raise the degree of disability by 7.5 percentage-points, on average, over low-waged applicants for whom there is any discrepancy between the final and potential degree of disability. The respective conditional mean difference for high-waged applicants is only 0.86 percentage-points. In the proximity to each of the other thresholds, the conditional mean difference between the final and potential degree of disability falls monotonically in going from low- to middle- to high-waged applicants. This pattern provides further evidence of greater discretion being exercised in determining the DI entitlement of the lower-waged applicants.

TABLE D4. DISCREPANCY BETWEEN FINAL DEGREE OF DISABILITY AND POTENTIAL DEGREE OF DISABILITY IF JOB MATCHES WERE NOT RULED INFEASIBLE – COVARIATE ADJUSTED

| | Degree of disability threshold | | | | |
|---|---|---|---|---|---|
| | 35% | 45% | 55% | 65% | 80% |
| Panel A: Proportion with final degree of disability > potential degree of disability | | | | | |
| Low wage | 0.422 | 0.483 | 0.412 | 0.525 | 0.347 |
| | (0.017) | (0.016) | (0.017) | (0.032) | (0.077) |
| Middle wage | 0.337 | 0.362 | 0.333 | 0.365 | 0.409 |
| | (0.010) | (0.013) | (0.011) | (0.011) | (0.045) |
| High wage | 0.298 | 0.310 | 0.334 | 0.348 | 0.376 |
| | (0.006) | (0.007) | (0.010) | (0.009) | (0.011) |
| N | 9,821 | 7,812 | 6,593 | 6,339 | 2,647 |
| B: Mean (final degree of disability – potential degree of disability \| final DD > potential DD) | | | | | |
| Low wage | 7.184 | 7.185 | 3.710 | 8.353 | 4.393 |
| | (0.363) | (0.415) | (0.445) | (0.800) | (2.430) |
| Middle wage | 1.925 | 2.241 | 2.361 | 2.895 | 7.910 |
| | (0.211) | (0.374) | (0.351) | (0.310) | (1.483) |
| High wage | 0.908 | 1.079 | 1.972 | 1.573 | 1.660 |
| | (0.133) | (0.231) | (0.301) | (0.264) | (0.308) |
| N | 3,144 | 2,742 | 2,292 | 2,304 | 1,002 |
| C: Proportion of those at or above threshold who would be below if used potential degree of disability | | | | | |
| Low wage | 0.374 | 0.339 | 0.171 | 0.243 | 0.339 |
| | (0.016) | (0.017) | (0.019) | (0.035) | (0.089) |
| Middle wage | 0.086 | 0.145 | 0.127 | 0.145 | 0.266 |
| | (0.009) | (0.014) | (0.013) | (0.012) | (0.051) |
| High wage | 0.058 | 0.047 | 0.088 | 0.109 | 0.149 |
| | (0.005) | (0.007) | (0.010) | (0.010) | (0.011) |
| N | 5,977 | 4,536 | 3,771 | 3,520 | 1,670 |

*Note*: Notes to Table C3 apply. That table shows the raw means of each outcome. This tables shows means adjusted for differences in the values of covariates across the three wage tercile groups. The adjusted means are obtained by estimating the following model: $Y_i = \beta_0 + \sum_{t=2}^{3} W_{it} + \beta_1 Age_i + \beta_2 FML_i + \beta_3 Male_i + E_i + \theta_i + \zeta_i + T_i + \delta_i + \varepsilon_i$, where $Y_i$ is the respective outcome for panels A, B, and C, $W_{it}$ are binary indicators of the bottom, middle, and top third of the pre-disability wage distribution, $Age_i$ is measured in completed years, $FML_i$ is the functional limitations score, $Male_i$ is an indicator of sex, $E_i$ are educational category fixed effects, $\theta_i$ are industry fixed effects, $\zeta_i$ are zip code fixed effects, $T_i$ are year of application fixed effects, and $\delta_i$ are main diagnosis fixed effects, and $\varepsilon_i$ is an error term. The models are estimated by ordinary least squares. For each wage tercile group, we predict the outcome if each observation were in that group and average these predictions over the sample used to estimate the model. The table shows these sample-averaged predictions for each group. Sample sizes are smaller than in Table C3 because of missing values on some covariates.

Panel C shows the proportion of applicants with a final degree of disability that places them at or above (but within 2 pp of) the respective threshold who would have been below this threshold, and so received a lower benefit, if the potential degree of disability had been used to determine their entitlement. For example, almost 40% of low-waged applicants who were just above the 35% threshold would have been below it if higher-paying job matches found by the algorithm had not been ruled infeasible. In comparison, not even 6% of similarly positioned high-waged applicants would have been denied the partial DI benefit awarded if the potential degree of disability had been used. At each of the other thresholds, lower-waged applicants were more likely to be lifted above the threshold, and so receive higher DI benefits, as a result of job matches being ruled infeasible.

The pattern observed in all three panels of Table D3 – lower-waged applicants benefit more from assessors ruling higher-paying job matches infeasible – is robust to conditioning on the functional limitations score, demographics, education, industry, diagnosis, and zip code fixed effects (Table D4).

To further test for selective exercise of discretion, we stratify by pre-disability wage tercile group and, for each assessor, calculate the proportion of all marginal applications for which the final degree of disability is higher than the degree that would have been obtained if the three highest-paying selected job matches had been used. Marginal applications are those given a final degree of disability within a five-percentage point (pp) bandwidth of any one of the entitlement thresholds. We use a wider bandwidth for this analysis to obtain enough applications handled by each assessor. For the same reason, we restrict the sample to assessors who handled at least ten applications within each bandwidth and at least 50 applications in total within each wage tercile group.

Figure D2 plots the quantile function of the low wage – high wage difference in the proportion of marginal applications in which discretion is exercised, i.e., final DD > DD if used three highest paying selected job matches. This shows that assessors who exercise discretion more often when evaluating low-waged (marginal) applicants are much more plentiful than assessors who resort to discretion more frequently with high-waged (marginal) applicants. Approximately 60% of the assessors exercise their discretion more frequently with low-waged applicants, while less than 35% do so more frequently for high-waged applicants. Furthermore, for one third of assessors, the probability of exercising discretion when evaluating low-waged applicants is at least 20 pp higher than the respective probability when evaluating high-waged applicants. Only for 6% of assessors is the probability of exercising discretion in the evaluation of high-waged applicants at least 20 pp higher than the respective probability for low-waged applicants.

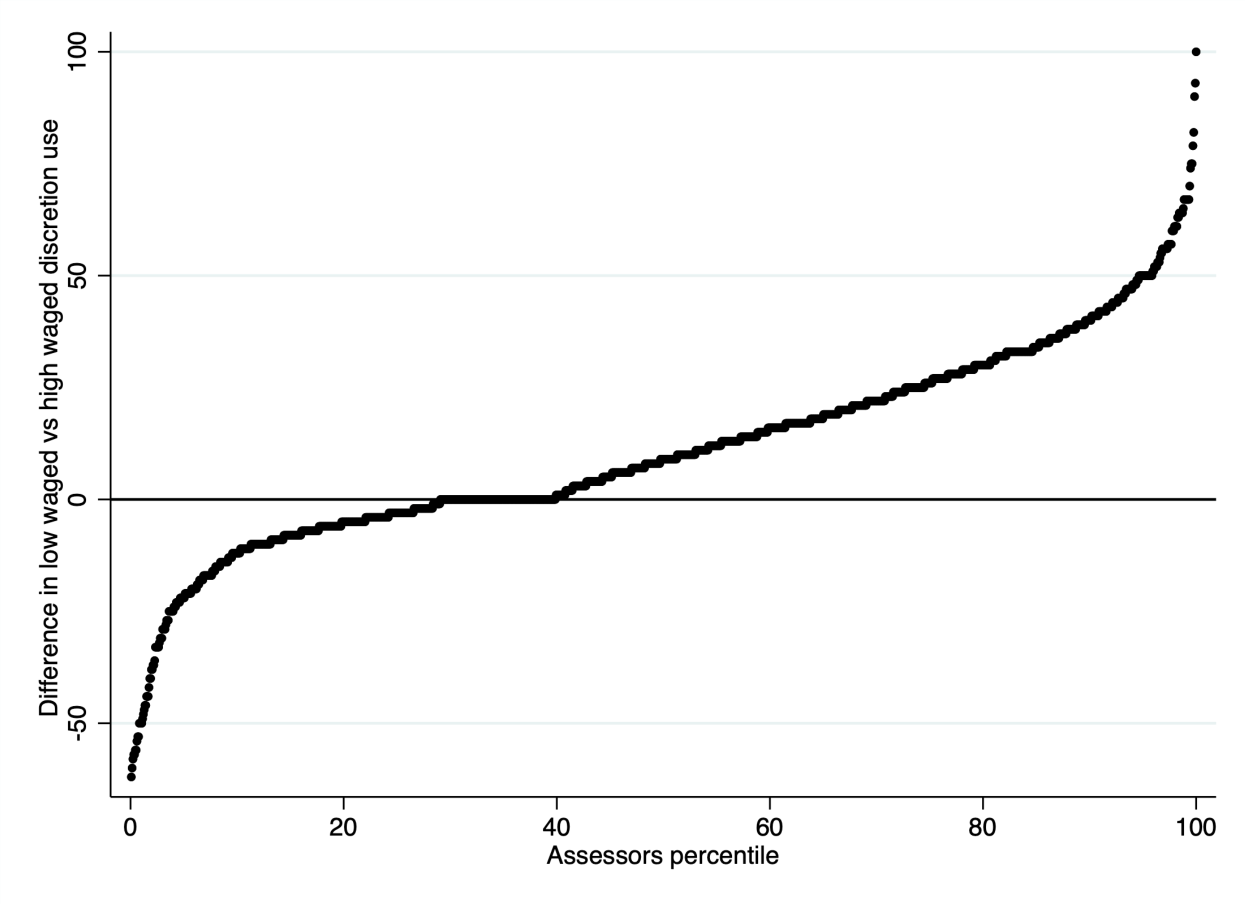


FIGURE D2. QUANTILE FUNCTION OF LOW-WAGE – HGH WAGE DIFFERENCE IN PROPORTION OF MARGINAL APPLICATIONS FOR WHICH DISCRETION EXERCISED

*Notes*: Plot of quantiles of assessor-specific low wage – high wage difference in proportion of applications within 5 pp of each DI entitlement threshold for which the final degree of disability is greater than the degree of disability that would arise if the assessor had used the three highest paying selected job matches. Positive values on the y-axis correspond to assessors who use discretion more often for low-waged applicants. Sample restricted to assessors who evaluated at least ten applications within each bandwidth and at least 50 applications in total within each wage tercile group. Sample includes 1,327 assessors and 84,744 applications.