Online Appendix

A Figures

Figure A1: DI enrollment fractions in the Netherlands (1998-2020) by impairment type.

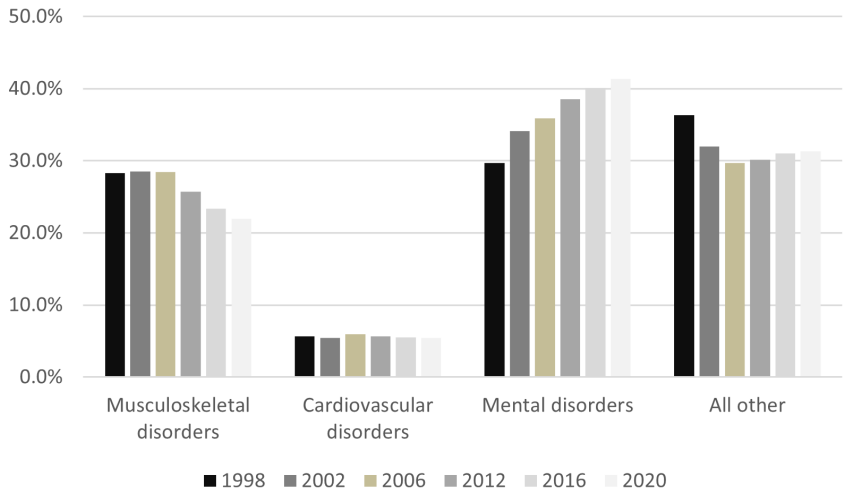
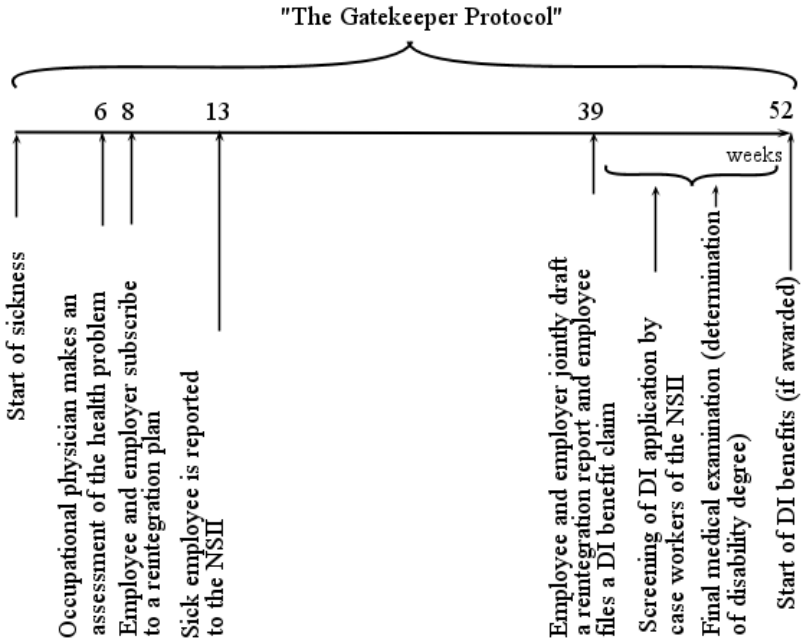


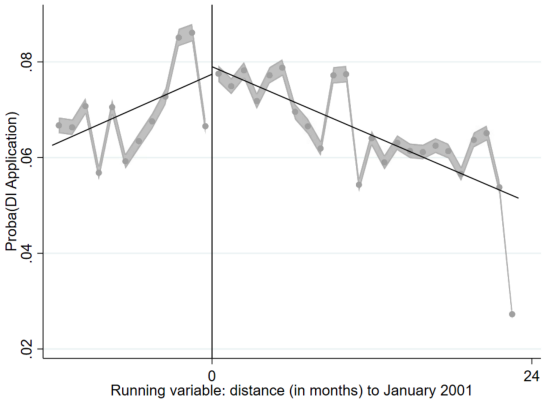
Figure A2: Schematic representation of the process toward entering DI



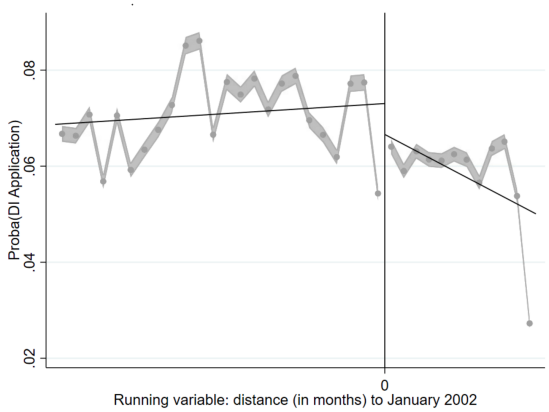
Note : Borrowed (and extended) from De Jong et al. (2011).

Figure A3: Placebo analysis: Regression-Discontinuity plots for DI application with cutoff at placebo dates.

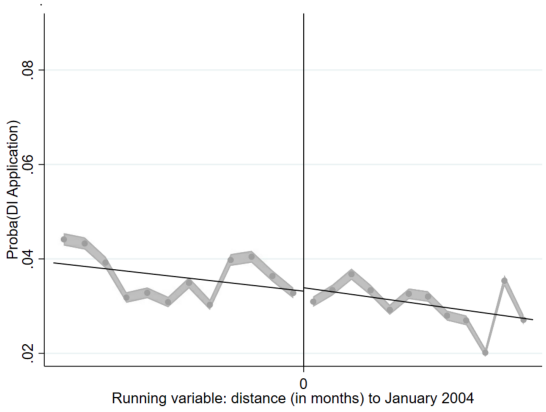
(a) Placebo cutoff: January 2001



(b) Placebo cutoff: January 2002



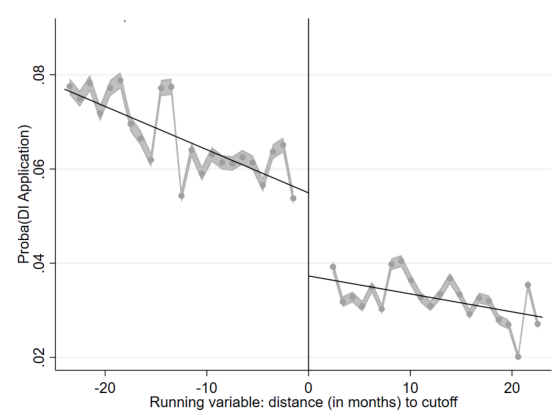
(c) Placebo cutoff: January 2004



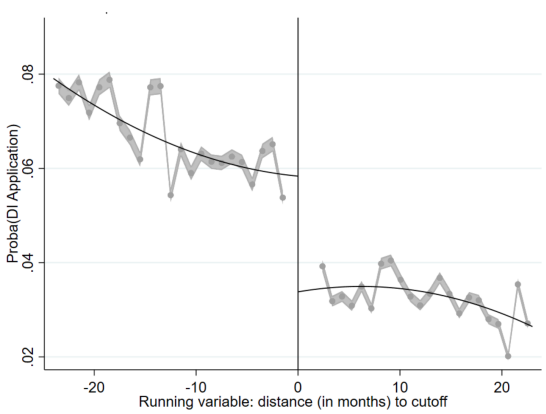
Note : The 95% confidence intervals around the binned means are calculated following Calonico et al. (2017). The data are binned by month, and a different standard deviation is estimated for each bin. The confidence interval for each bin is centered on the bin mean and based on that standard deviation.

Figure A4: Donut Regression-Discontinuity plots (dropping those within one month of the cutoff) for DI application with various choices of polynomial time control.

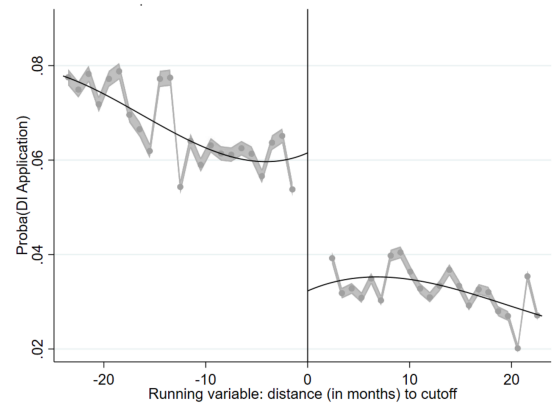
(a) Polynomial fit of order 1



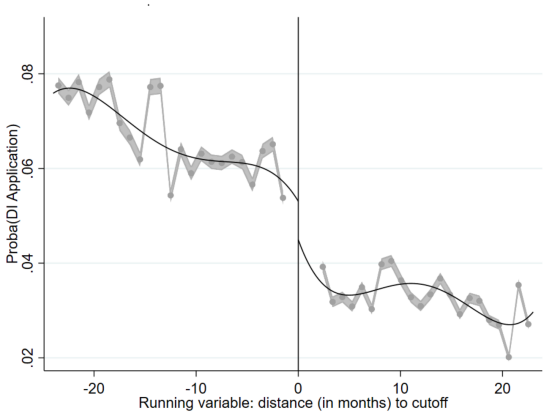
(b) Polynomial fit of order 2



(c) Polynomial fit of order 3

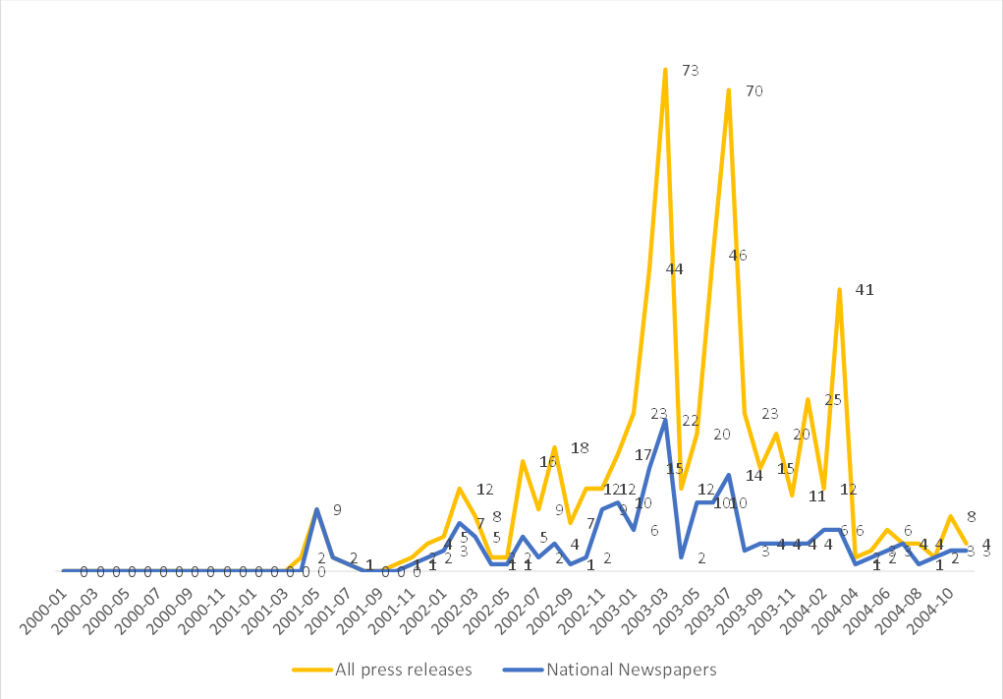


(d) Polynomial fit of order 4



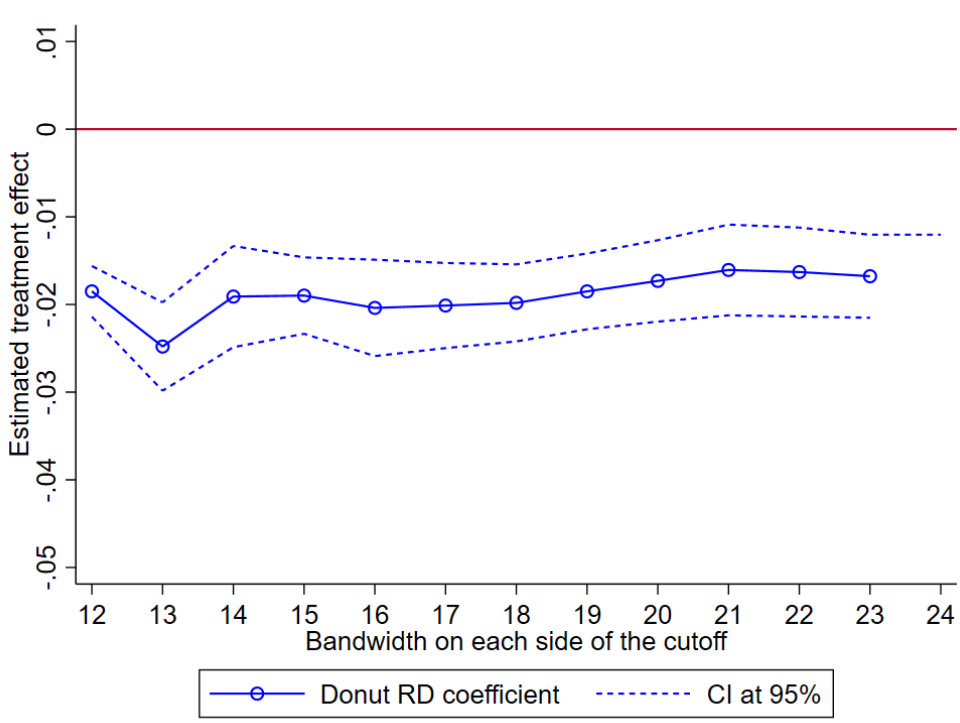
Notes : (i) Donut RD plots dropping individuals within one month of the cutoff. (ii) The 95% confidence intervals around the binned means are calculated following Calonico et al. (2017). The data are binned by month, and a different standard deviation is estimated for each bin. The confidence interval for each bin is centered on the bin mean and based on that standard deviation.

Figure A5: Press Coverage of Gatekeeper Protocol (2000-2004)



Note : Article numbers referring to the Gatekeeper Protocol are derived from the LexisNexis Academic database. This database includes all newspaper and magazine articles since 2000.

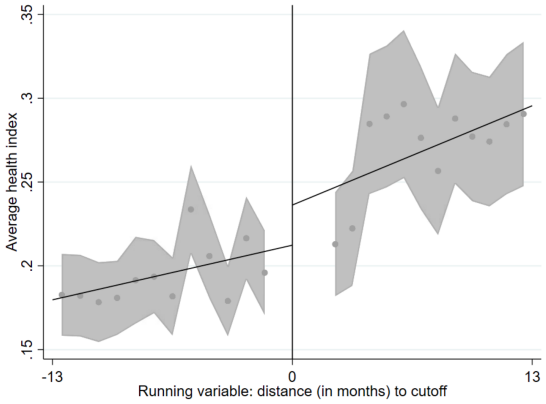
Figure A6: The effect of the Gatekeeper Protocol on DI application behaviour. Donut Regression-Discontinuity estimates (dropping those within one month of the cutoff) with varying bandwidths.



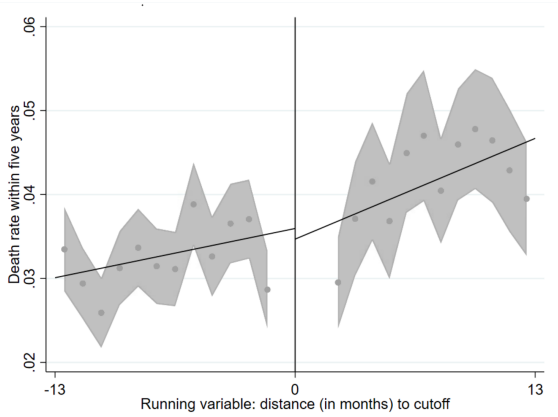
Note: Estimates are obtained by running Equation (1). Note that in the presence of month-of-year fixed effects, the model is not reliable for bandwidths below 12 months on each side of the threshold.

Figure A7: Donut Regression-Discontinuity plots (dropping those within one month of the cutoff) for **applicants** characteristics, 1/4.

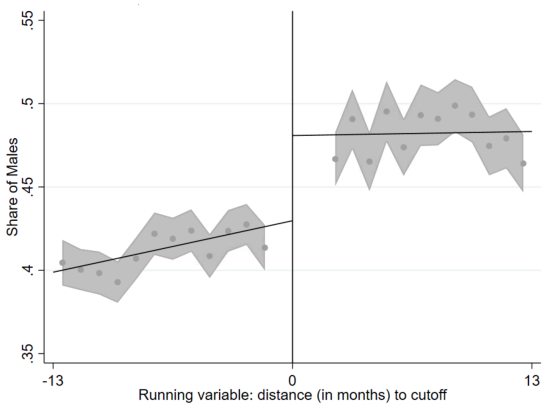
(a) Health Index



(b) Death rate within five years



(c) Share of males



(d) Share of natives

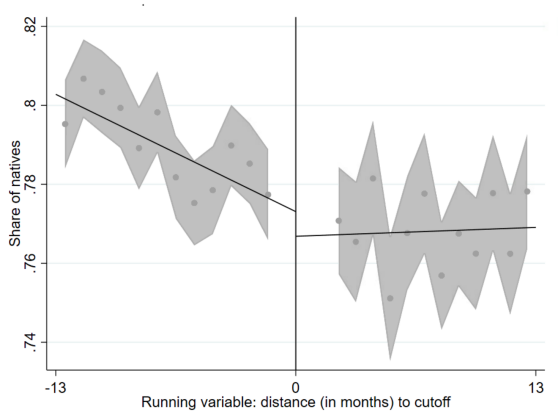
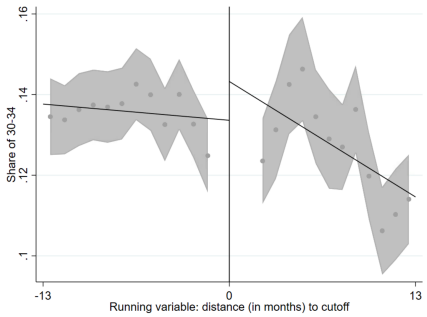


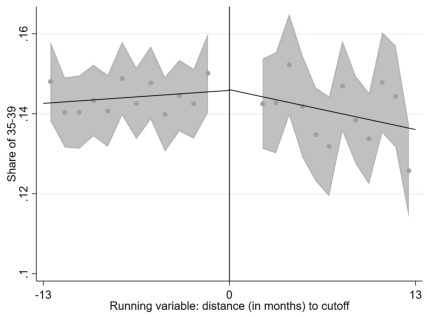
Figure A7: Donut Regression-Discontinuity plots (dropping those within one month of the cutoff) for **applicants** characteristics, 2/4.

(e) Share of applicants aged 25-29

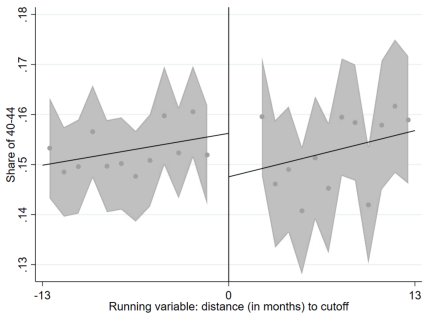
(f) Share of applicants aged 30-34



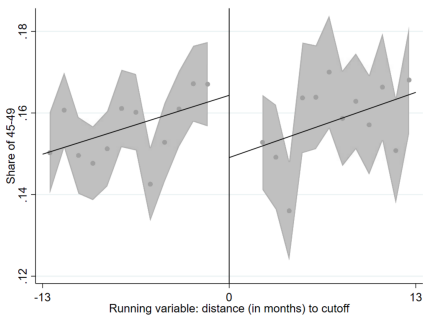
(g) Share of applicants aged 35-39



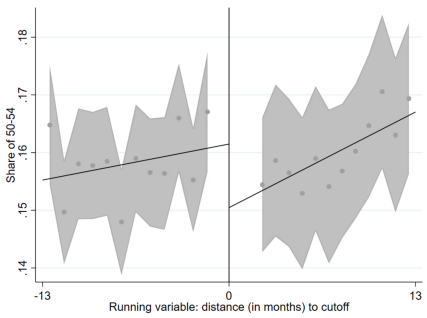
(h) Share of applicants aged 40-44



(i) Share of applicants aged 45-49



(j) Share of applicants aged 50-54



(k) Share of applicants aged 55-59

(l) Share of applicants aged 60-65

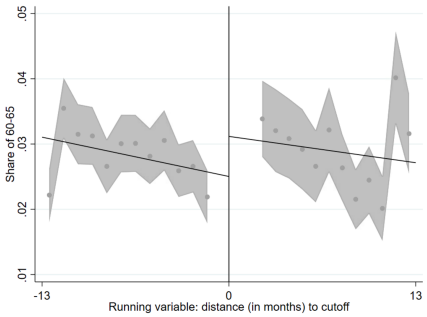
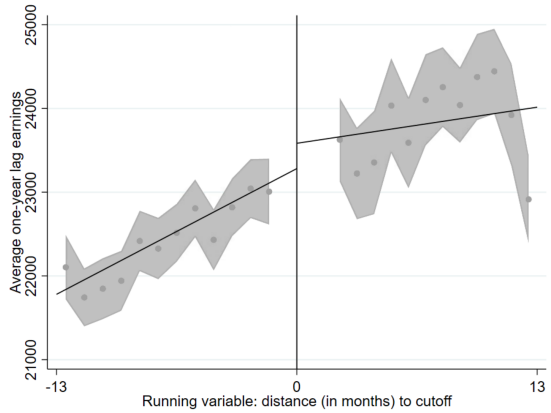
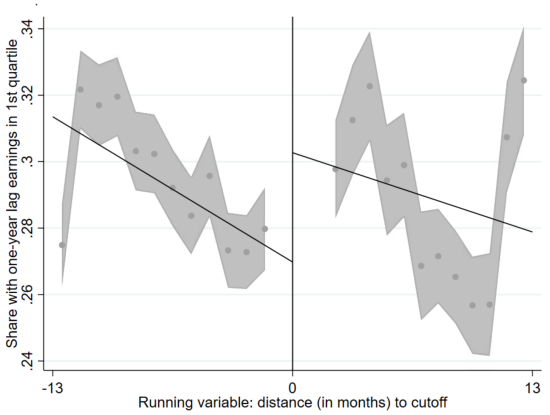


Figure A7: Donut Regression-Discontinuity plots (dropping those within one month of the cutoff) for **applicants** characteristics, 3/4.

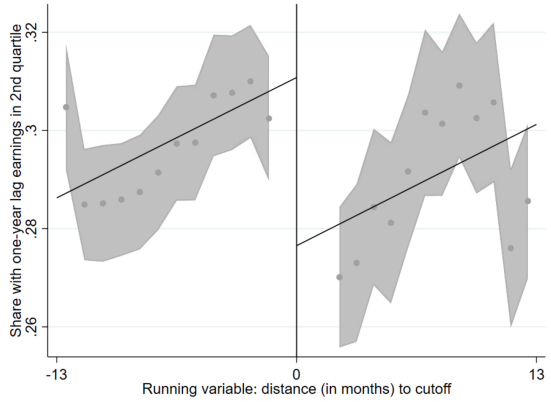
(m) One-year lag earnings (continuous variable)



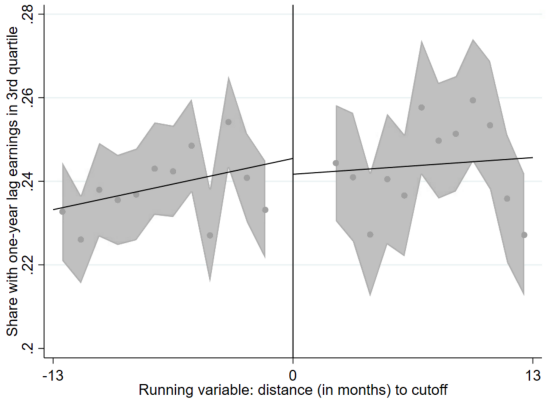
(n) Share of applicants in first quartile of the (one-year lag) earnings distribution.



(o) Share of applicants in second quartile of the (one-year lag) earnings distribution.



(p) Share of applicants in third quartile of the (one-year lag) earnings distribution.



(q) Share of applicants in fourth quartile of the (one-year lag) earnings distribution.

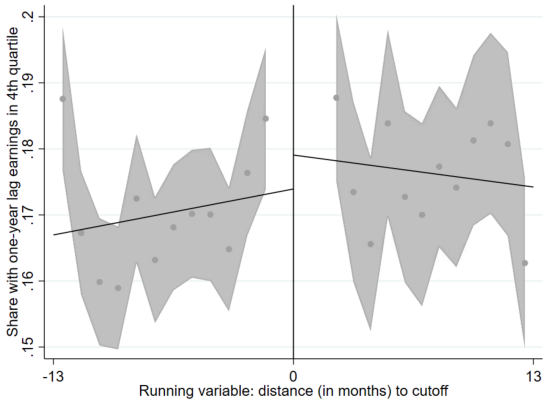
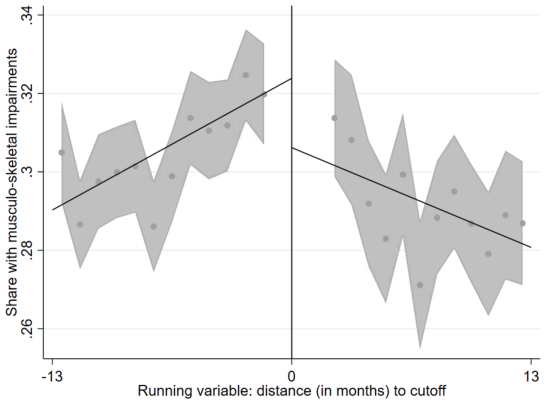
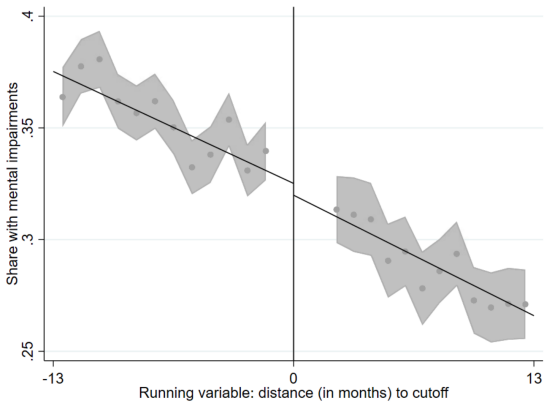


Figure A7: Donut Regression-Discontinuity plots (dropping those within one month of the cutoff) for **applicants** characteristics, 4/4.

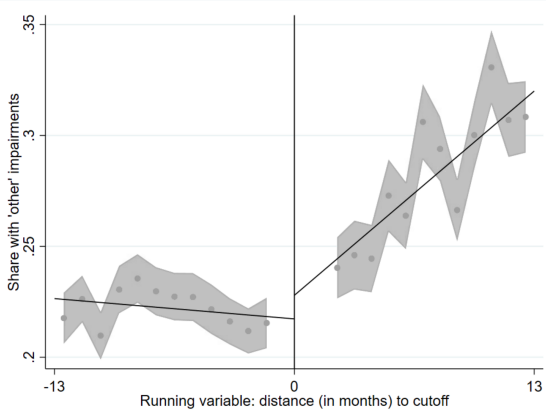
(r) Share applying for musculo-skeletal impairments



(s) Share applying for mental impairments



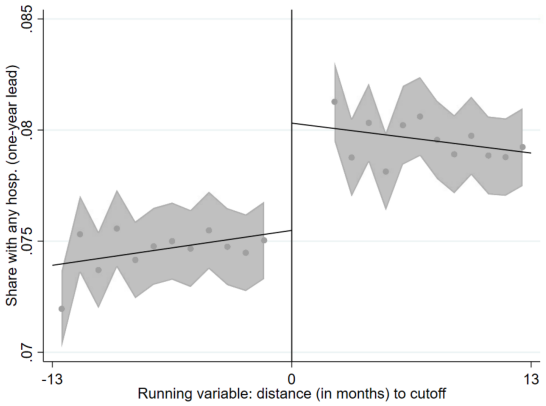
(t) Share applying for "other" (difficult-to-verify) impairments



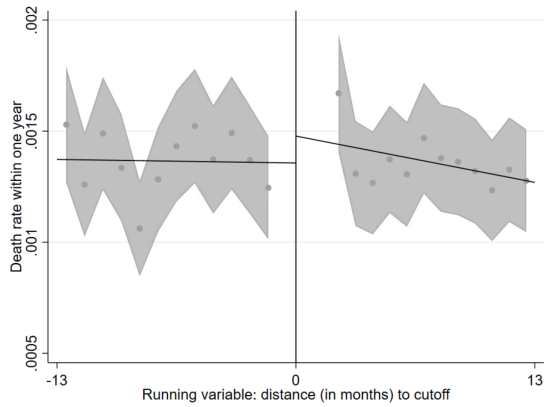
Note : The 95% confidence intervals around the binned means are calculated following Calonico et al. (2017). The data are binned by month, and a different standard deviation is estimated for each bin. The confidence interval for each bin is centered on the bin mean and based on that standard deviation.

Figure A8: Donut Regression-Discontinuity plots (dropping those within one month of the cutoff) for future (one-year lead) **non-applicants** outcomes, 1/2.

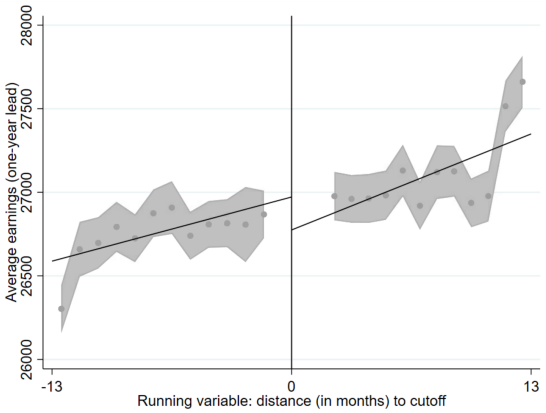
(a) Any hospitalization



(b) Death rate within five years



(c) Earnings (gross, 2010 euros)



(d) Share in employment

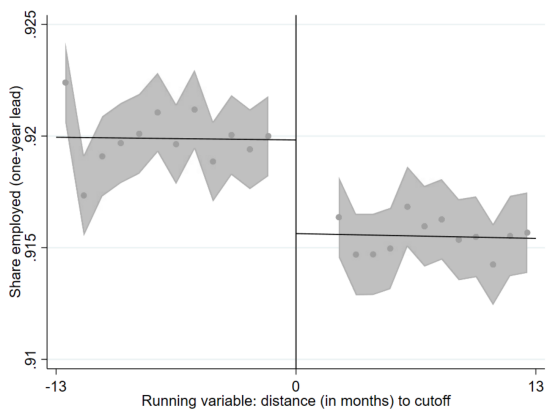
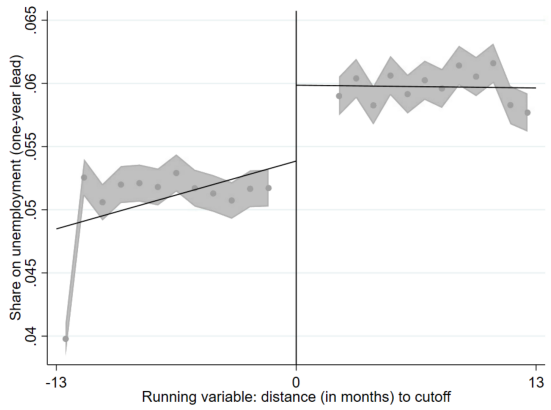
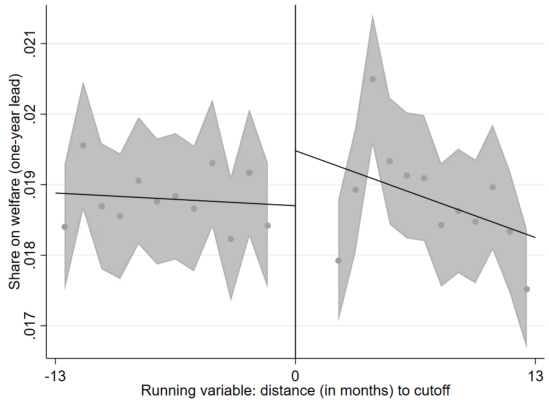


Figure A8: Donut Regression-Discontinuity plots (dropping those within one month of the cutoff) for future (one-year lead) **non-applicants** outcomes, 2/2.

(e) Share receiving unemployment benefits

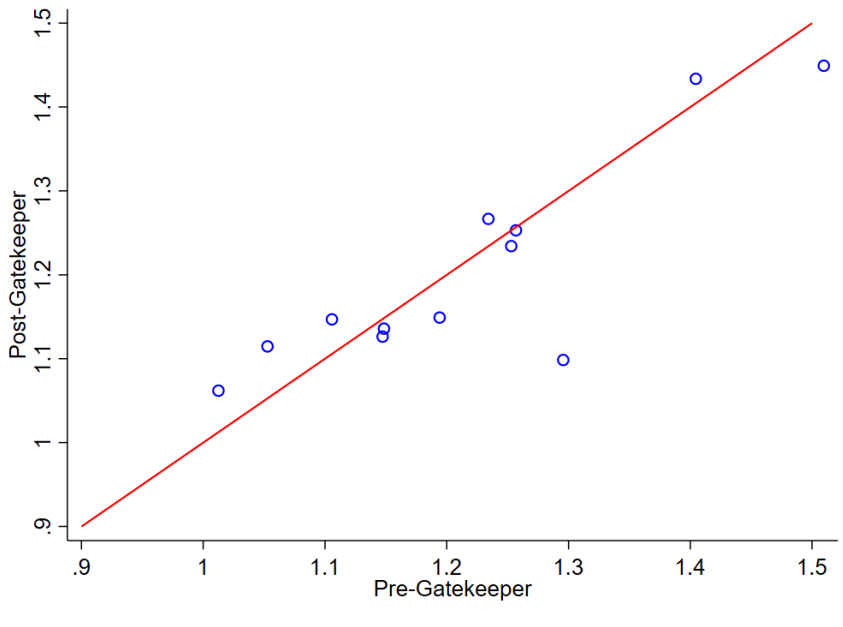


(f) Share receiving welfare benefits



Note : The 95% confidence intervals around the binned means are calculated following Calonico et al. (2017). The data are binned by month, and a different standard deviation is estimated for each bin. The confidence interval for each bin is centered on the bin mean and based on that standard deviation.

Figure A9: Relative award rates (δ), before and after the introduction of the Gatekeeper protocol, for subgroups of applicants.



Note : Subgroups of applicants are defined according to gender, age groups (young/primeage/senior) and impairment types (difficult/easy-to-verify). See Section 4.3 for more details about the calculation of δ, which can be expressed in terms of expected health values of applicants and awarded applicants. We use the Charlson health index as a proxy for health.

B Tables

Table B1: Summary statistics of the data[(a)].

	Pre-Gatekeeper 2001-2002 (1)	Post-Gatekeeper 2003-2004 (2)
Demographics		
Age	38.69	39.19
	(9.56)	(9.58)
Male	0.56	0.55
	(0.50)	(0.50)
Native	0.82	0.82
	(0.38)	(0.39)
Labour-market[(b)]		
Employed	0.95	0.95
	(0.21)	(0.23)
Gross earnings (2010 euros)	25,829	27,108
	(22,117)	(23,613)
UI recipient	0.04	0.06
	(0.19)	(0.23)
Welfare recipient	0.02	0.02
	(0.14)	(0.13)
DI		
DI (monthly) application rate	0.065	0.033
	(0.025)	(0.018)
DI (monthly) award rate (uncond.)	0.03	0.02
	(0.019)	(0.014)
Health status		
Health index	0.036	0.040
	(0.373)	(0.398)
Hospitalized in the three previous years[(c)]	0.168	0.168
	(0.374)	(0.374)
Hospitalization type (among hospitalized)[(d)]		
Musculo-skeletal disorders	0.180	0.174
	(0.384)	(0.379)
Neoplams	0.060	0.060
	(0.237)	(0.237)
Cardiovascular diseases	0.088	0.088
	(0.284)	(0.283)
Mental disorders	0.010	0.009
	(0.101)	(0.094)
Endocrine problems	0.013	0.014
	(0.115)	(0.116)
Nervous disorders	0.056	0.059
	(0.231)	(0.236)
Dead within five years	0.009	0.009
	(0.096)	(0.093)
Number of observations	2,240,348	2,261,923

Notes : [(a)] For each month in 2001–2004, the sample includes all individuals for whom we observe an award decision that month ('applicants') as well as a 1% random sample of the rest of the population ('non-applicants') – see Section 3. [(b)] Our sample excludes individuals not employed in the previous year (see Section 3). Labour-market characteristics are measured at the yearly level; [(c)] For each individual, we construct indicators for whether an individual was hospitalized (all-cause or cause-specific) in t-1; t-2 or t-3; [(d)] Note that only the main hospitalization types are listed here. Individuals can be hospitalized for several reasons in the previous three years, so lines could add up to more than 100%.

Table B2: The effect of the Gatekeeper Protocol on future outcomes (one-year-lead) of **non-applicants** – Donut RD estimates.

Future (one-year lead):	**Change in mean [% change]**	**Implied difference "leavers" vs. non-participants**
Panel A: Health and Mortality		
Hospitalization (any type)	0.004*** [+5.9%]	0.071***
	(0.001)	(0.010)
Death rate	0.0002** [+16.2%]	0.007***
	(0.0001)	(0.003)
Panel B: Labor-market outcomes		
Earnings (gross, 2010 euros)	-438*** [-1.6%]	-9,442***
	(54)	(896)
Employment (fraction)	-0.004*** [-0.4%]	-0.086***
	(0.001)	(0.023)
UI receipt (fraction)	0.005 [+9.5%]	0.086***
	(0.003)	(0.021)
Welfare receipt (fraction)	0.001* [+5.0%]	-0.003
	(0.000)	(0.006)
Number of individuals	2,150,213	2,150,213

Notes : (1) Each line in column (i) presents the estimated impact of the GKP reform on the average value of a characteristic (e.g. average health index) or the proportion of non-applicants with a given characteristic. (2) Donut RD dropping those within one month of the cutoff. We use a bandwidth of 13 months on each side of the cutoff and include a linear trend that is flexible on either side of the cutoff, as well as month-of-year dummies (see Equation 1). (3) Standard errors in column (i) are clustered both at the individual level and at the month-of-year level. Standard errors in column (ii) are obtained from seemingly unrelated regression, and clustered at the month-of-year level. (4) Power calculations show that restricting our sample to a 1% random sample of all non-applicants is sufficiently large to detect reasonably sized effects for most outcomes in column (1). The only exception is for UI receipt, for which we would need a 10% random sample of non-applicants to detect a significant effect at the 5% level and a power of 20%. (5) *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

C Modeling the effect of stricter screening on applications

Stricter screening of DI applications under the Gatekeeper Protocol (GKP) may affect application behavior of disabled and able workers in two ways: (i) increased application costs and (ii) decreased noise of the disability signal of applications. On the one hand, the protocol requires more tests, more reintegration efforts and more documents have to be provided to support the DI claim. This will increase DI application costs with effects that may vary according to disability status. For disabled workers the costs could even be larger than for able workers, because of higher mortality expectations, or because these reintegration efforts are more stressful and burdensome than for able workers (Parsons, 1991). On the other hand, increased screening leads to more information about the true disability status of the applicant. This means that for disabled workers acceptance rates may increase, while for able workers they may decline. The type of impairment may also affect costs and benefits. Marginal changes in costs and acceptance rates will be smaller for impairments that are easy–to–verify than for impairments that are more difficult–to–verify. For instance, for easy verifiable impairments, like last stage cancers, the prospects of return to work may be very low and no (additional) reintegration efforts are needed. For diseases that are more difficult to verify, such as mental impairments, more tests may be required and the effectiveness of different interventions may not always be clear.

Below we formulate a simple model of worker application behavior that captures these above-mentioned features. Like Parsons (1991) we assume that there are two types of individuals:

Type 0, the Disabled, who are unable to work, and Type 1, the Able, who can work. Those in work receive a wage W. DI benefits are equal to B and other non-work states provide social assistance benefits SA, with $SA < B < W$. Those who report sick enter a waiting period T that precedes the DI application. Unlike e.g. the US where individuals receive no income during the waiting period, Dutch workers receive sick pay up to 100% of their net earnings. During this waiting period sick workers have to follow the GKP protocol that requires medical checks and work resumptions efforts of the worker (and employer). Costs C associated with these medical checks and work resumption efforts differ by worker type j, for $j = 0, 1$. For instance, this may capture the fact that the hassle and efforts are more taxing for those in bad health. Application costs are increasing in screening intensity s : $C_j(s)$, $C_j'(s) \geq 0$, for $j = 0, 1$. The GKP increased screening scrutiny and hence increased application costs.

Award decisions ϕ differ by worker type and depend on screening scrutiny s: $\phi_j(s)$, for $j = 0, 1$. More scrutiny in the screening process increases the award rate for disabled workers (Type 0) and reduces the award rates of able workers (Type 1): $\phi_j'(s) \geq 0$, for $j = 0$ and $\phi_j'(s) \leq 0$, for $j = 1$. An individual applies for a DI benefit if the expected benefits exceed expected application costs:

$$I_j = E\Big[U_j(\text{apply})\Big] \;>\; E\Big[U_j(\text{not apply})\Big] \quad \text{for} \quad j = 0, 1$$

The expected utility of an application for a *disabled* worker consists of two parts. For the waiting period, this is the net discounted value of the utility from wages (workers receive sick-pay up to 100% of their net earnings) and the disutility from the application costs. Note that the discount rate for disabled is defined as ρ_0. After the application, we have two possible

outcomes with corresponding utility levels: acceptance with income from DI benefits (B) or rejection with income from social assistance benefits (SA):

$$E\Big[\, U_0(\text{apply}) \,\Big] = \int_0^T \Big[U(W) - C_0(s)\Big] e^{-\rho_0 t} dt \; + \\ + \; \phi_0(s) \int_T^\infty U(B) e^{-\rho_0 t} dt + \Big(1 - \phi_0(s)\Big) \int_T^\infty U(SA) e^{-\rho_0 t} dt \tag{8}$$

Since Type 0 applicants are unable to work, the expected utility of not applying equals the discounted value of utility from SA:

$$E\Big[\, U_0(\text{not apply}) \,\Big] = \int_0^\infty U(SA) e^{-\rho_0 t} dt \;=\; \frac{U(SA)}{\rho_0} \tag{9}$$

We next calculate the derivative of the net expected discounted utility of applications with respect to the screening intensity s:

$$\frac{\partial I_0}{\partial s} \;=\; \frac{1}{\rho_0} \left[\frac{\partial C_0(s)}{\partial s} (e^{-\rho_0 T} - 1) \;+\; \frac{\partial \phi_0(s)}{\partial s} \Big[U(B) - U(SA) e^{-\rho_0 T} \Big] \right] \tag{10}$$

The sign of Equation (10) is undetermined for disabled workers. Increases in screening stringency will result in more applications if the positive effect of screening scrutiny on award rates (the second term on the RHS of Equation (10)) outweighs the negative effect of screening scrutiny on costs (the first term on the RHS of Equation (10)). Disabled workers may also screen themselves out if screening scrutiny increases costs more than award rates. This may in particular be relevant for diseases that are more difficult to verify, such as mental disorders that require more medical tests and where the effectiveness of interventions are more difficult to assess. Such self-screening effects are also more likely to occur when disabled workers perceive to have worse mortality expectations (i.e. when ρ_0 is large), which is relevant for certain conditions.

We next turn to the application decision of*able* workers. For this group, increases in screening stringency will always reduce DI application rates. Defining k as the disutility of work, the expected utility of applying for a DI benefit by an able worker again equals the net discounted value of the utility in the waiting period and the expected discounted value that follows after the award decision (with the award probability $\phi_1(s)$ and discount rate ρ_1):

$$E\Big[\, U_1(\text{apply})\,\Big] = \int_0^T \Big[U(W) - C_0(s)\Big] e^{-\rho_1 t} dt \; + \\ + \; \phi_1(s) \int_T^\infty U(B) e^{-\rho_1 t} dt + \Big(1 - \phi_1(s)\Big) \int_T^\infty \Big[U(W) - k\Big] e^{-\rho_1 t} dt \quad (11)$$

And the expected utility of not applying for Type 1 workers is

$$E[U_1\Big(\text{ not apply})\,\Big] \;=\; \int_0^\infty \Big[U(W) - k\Big] e^{-\rho_1 t} dt \;=\; \frac{U(W) - k}{\rho_1}. \quad (12)$$

The derivative of the net benefits of applications with respect to the screening intensity s for Type 1 workers thus equals

$$\frac{\partial I_1}{\partial s} \;=\; \frac{1}{\rho_1} \left[\frac{\partial C_1(s)}{\partial s} (e^{-\rho_1 T} - 1) \;+\; \frac{\partial \phi_s(s)}{\partial s} \Big[U(B) + k - U(W)\Big] e^{-\rho_1 T} \right], \quad (13)$$

which is always negative because both terms on the right hand side of Equation (13) are negative.[39] Increased application costs and the increased likelihood of rejections both decrease the option value of an application. This contracts to the effects for the disabled workers. From this alone, however, we cannot conclude that the decrease in applications will always be higher for Type 1 than for Type 0 workers. This depends on possible differences in discount rates and differences in the disutility inherent with increased application costs. For relatively high discount rates and high application cost, it may even occur that the increased scrutiny leads

[39]Note that a worker will only consider an application if $U(B) \geq U(W) - k$.

to perverse screening, i.e. that increased application costs disproportionately deter disabled workers from applying Parsons (1991)).

D Data Appendix : The Charlson Comorbidity Index

The Charlson Comorbidity Index (CCI) is a popular tool for predicting mortality by classifying or weighting comorbid conditions (comorbidities) – see Charlson et al. (1987). The CCI can be constructed from medical record abstract or administrative data. We use the coding algorithm developed by Stagg et al. (2015) to derive the CCI from ICD-9-CM administrative data. For each individual, in each year, we compute the CCI index as a weighted sum of 17 comorbidities. The 17 comorbidities and their associated weights – that allow for adjustment for severity of illness – are listed in Table D1. As we have longitudinal data, we then compute a time-varying comorbidity index that aggregates information over multiple hospitalization spells since 1995.

Table D1: Charlson Comorbidities and Weights

Comorbidity	**Assigned weight**
Acute Myocardial infection	1
Congestive Heart Failure	1
Peripheral Vascular Disease	1
Cerebrovascular Disease	1
Dementia	1
Chronic pulmonary disease	1
Rheumatic disease	1
Peptic Ulcer Disease	1
Mild Liver Disease	1
Diabetes without chronic complications	1
Diabetes with end organ damage	2
Hemiplegia / Paraplegia	2
Renal (kidney) Disease	2
Cancer (Any malignancy/lymphoma/leukemia)	2
Moderate or severe liver disease	3
Metastasic Cancer	6
AIDS/HIV	6