Targeting Disability Insurance Applications with Screening

Mathilde Godard

Pierre Koning

Maarten Lindeboom

January 24 2019, Dublin

Background

- Expenditures on DI schemes are growing to levels that are much higher than any other social insurance scheme
- Policy makers: reduce DI benefits or tighten eligibility conditions
- Reductions in benefits reduce costs, but harms truly disabled workers
- Calls for policies that improve the screening efficiency of DI application program
  - More rigor in screening/reintegration efforts etc
  - With the aim to improve Targeting Efficiency (TE)

- Such considerations also played a role in the Netherlands:

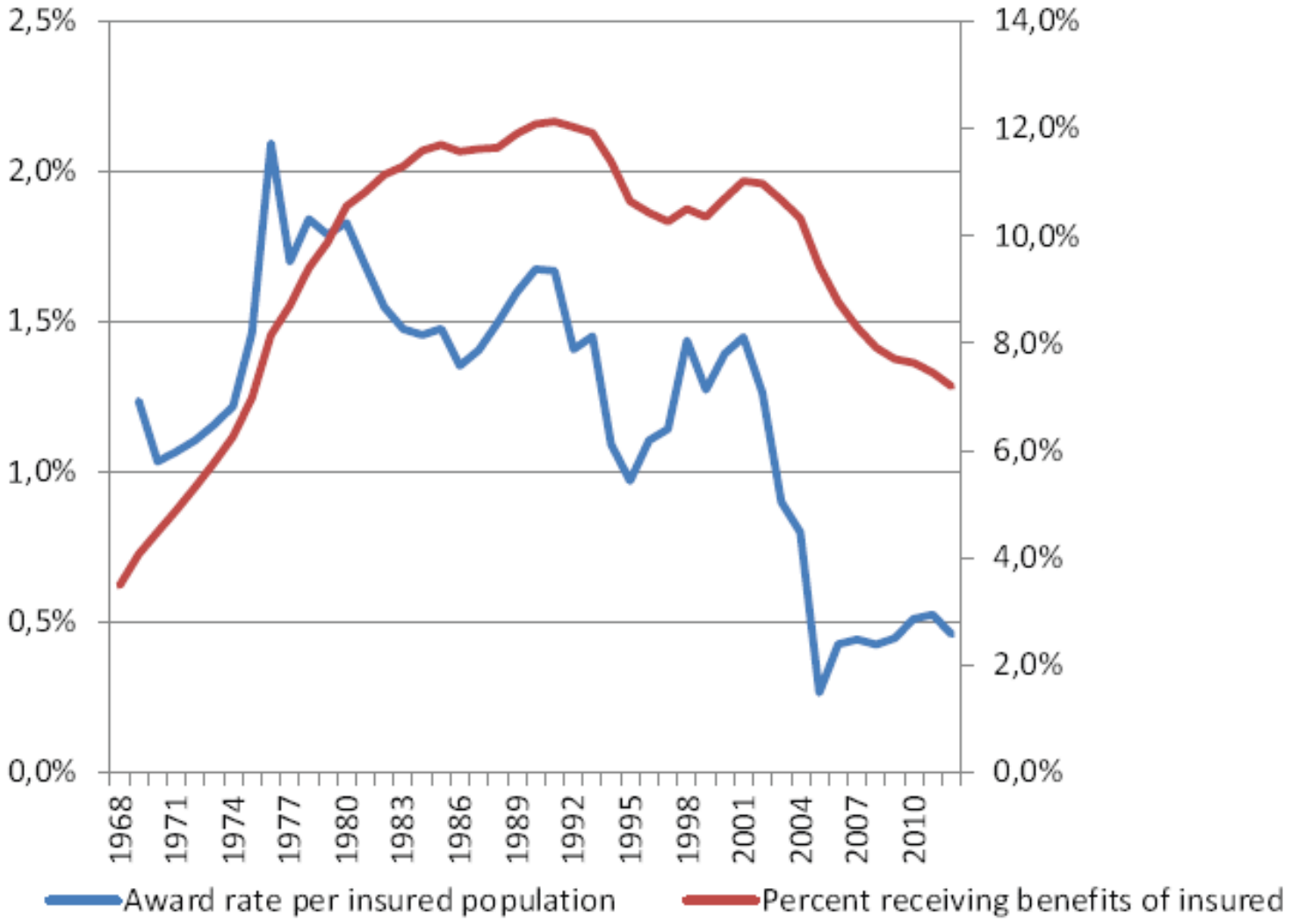
  "most out of control program of OECD countries"

- Despite series of reforms since late 1990's enrolment rates and Inflow rates remained high

  < see figure next slide >

- Substantive drop in figure coincides with implementation of a drastic reform: Gatekeeper protocol (2002)
- Central element in this reform is the screening of DI applicants & structured way to increase work resumption in waiting period

Figure 1: Disability Insurance Award and Enrolment Rate per Insured Worker in the Netherlands, 1968-2012



Objective of paper

1. Main goal of the reforms was to reduce DI inflow
2. To increase employment rates of workers with disabilities
3. Providing benefits to those who really need them (improving targeting efficiency)

- It seems that at least goal 1 was met, but less clear for other two goals
- Here: examine the effects of (intensified) screening on targeting efficiency of the program

- Disability is difficult to observe and costly to verify

  => classification errors (false positives/negatives)

- Primary system for targeting of DI benefits is the disability determination process
- National Social Insurance Institute (NSSI) determines whether individual meets eligibility criteria

- Screening is central in determination process
- Increases in rigor of screening during application phase may:
  - Reduce classification errors

    (False positives (Type II classification errors) and false rejections (Type I classification error)
  - But also increase costs (more test/reintegration etc) and thus affect decision of individuals to report sick and file DI claims
  - If such deterrence effects confined to healthiest

    => screening improves Targeting Efficiency
  - However, it may also disproportionally deter those in bad health to apply

    => Perverse self-screening (Parsons, 1981)

This paper

- We empirically assess targeting effects of (stricter) screening
  - Who is screened out (i.e. who stops applying)?
  - Look at pool of applicants and non-applicants
  - Look at award rate changes in final stage of application process
- To address this we exploit two sources of variation
  i. In time, implementation of reform; (extensive margin)
  ii. Field experiment (intensive margin of screening)
- And use rich Dutch administrative data covering the period 2001 - 2008
  - Hospitalizations, mortality and labor market outcomes

Relevant literature

- Literature on effect of benefits/eligibility/denial rates public DI programs
  - Staubli, 2011, Borghans et al, 2014, Moore 2015, Karlstrom et al, 2008, Autor et al, 2015)
- Literature on link between imperfect information about work disabilities and classification errors
  - Akerlof, 1978, Parsons, 1996, Kleven & Kopzcuk,2011
- Few papers on effect of application costs on DI applications
  - Maestas et al, 2013, Autor et al, 2015, Deshpande & Li, 2017, Markussen at al, 2017.

Our paper

i. Assess targeting efficiency (TE) screening return-to-work efforts during application process
   - Self-screening: who is screened out?
   - Composition effects in pool of (non-)applicants

ii. Are there (further) improvements in TE at the final stage of the application process (by medical examiners)

iii. Look at extensive margin (effect of implementation of Gatekeeper protocol) and intensive margin (exploit field experiment that changed intensity of screening)

iv. Use rich (linked) adminstrative data bases covering the full dutch population over 2001-2008)

DI process in more detail

- All disabilities (in or out of work) are insured
- Replacement rate during waiting period effectively 100%
- Employer has to contract occupational health services
- Degree of disablement depends on loss in earnings capacity
- Once application approved: Wage related Benefits of 70% of last gross wage
- Time of wage related benefits depends on age of onset
- At age 58 effectively retired

The gatekeepers protocol

- Responsibility of reintegration of sick worker in waiting period left to worker and employer
- National Social Insurance Institute (NSII) follows gatekeeper protocol (GKP) for DI application process
- Structured process, where employer bears most costs

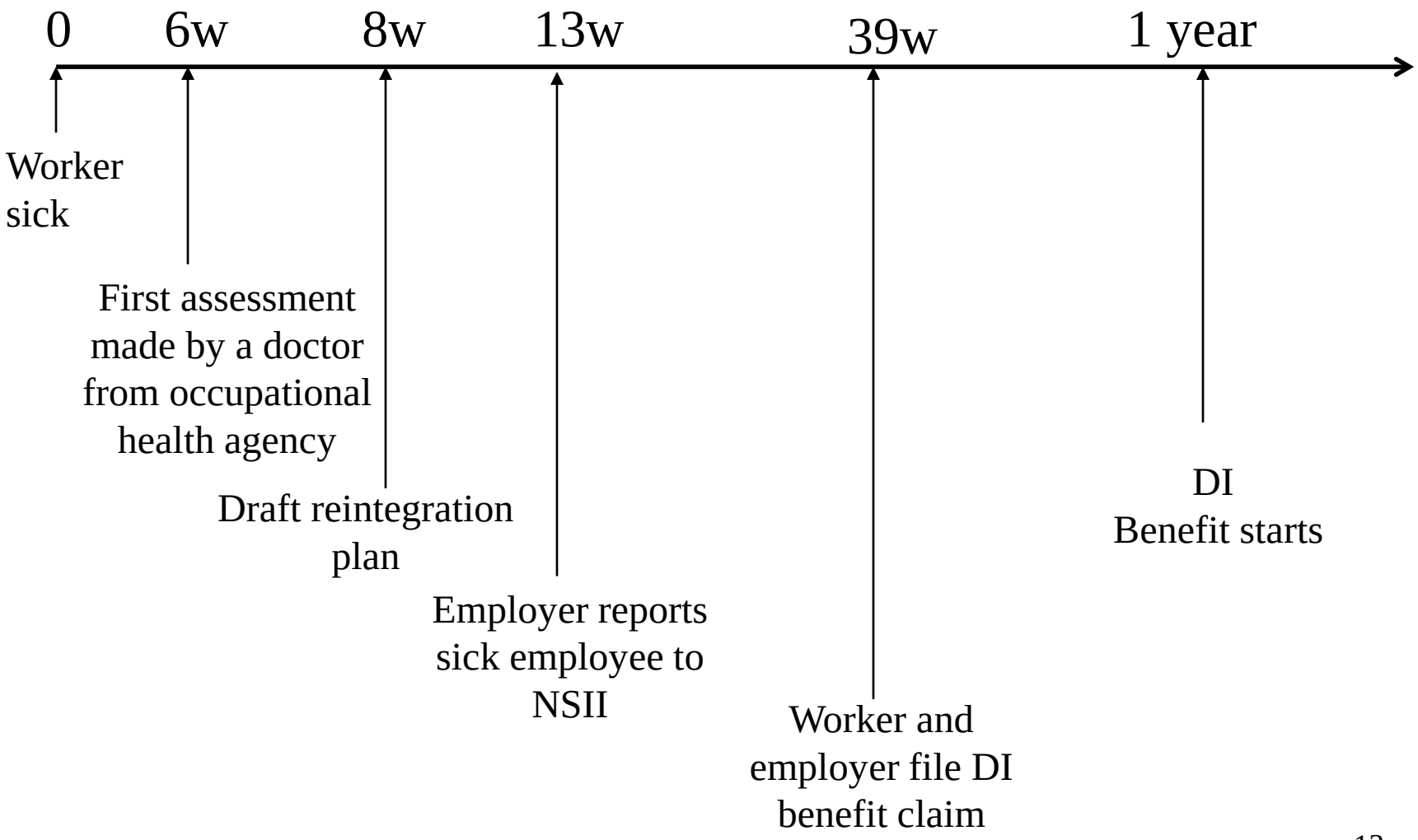
< see next slide>

→ Central element is screening of reintegration report

→ Insufficient effort can lead to a sanction for employer

(Employer can fire worker who does not comply)

Schematic representation of gatekeeper protocol



- Gatekeeper screening is implemented nationally and applies to all workers becoming sick after April 2002.
- DI applications are made in week 39 => applications affected as of January 2003

To structure thoughts a little

<u>The employer side</u>

- DI application process is very costly for Employers
  - → Financially responsible for sick pay during waiting period and experience rating in DI
  - → Mandated to contract occupational health agencies
- GKP: Employer is fully responsible in case of non-compliance
- So GKP forces employers to increase reintegration efforts to at least a minimum requirement level

  => Work resumption rates↑ => # DI applications ↓
- Effect may differ by disease type
  - — Difficult to verify versus Easy to verify

The worker side

- Key question, whether GKP influenced decision to enter application process.
- Two ways this decision is affected
  a) Costs ↑ (more intervention/verifcation/doctor visits)
  b) Decreases in the noise of the disability signal => gatekeeper knows more => classification errors ↓

- b) most relevant for worker with difficult to verify condition, in particular the most able worker in this group may be deterred from applying
- Those in this group with severe disabilities may benefit from more info (and may thus increase applications)
- However, costs may deter these deserviong worker from applying (the risk averse, those sentivive to extra efforts, low mortality expectations etc)

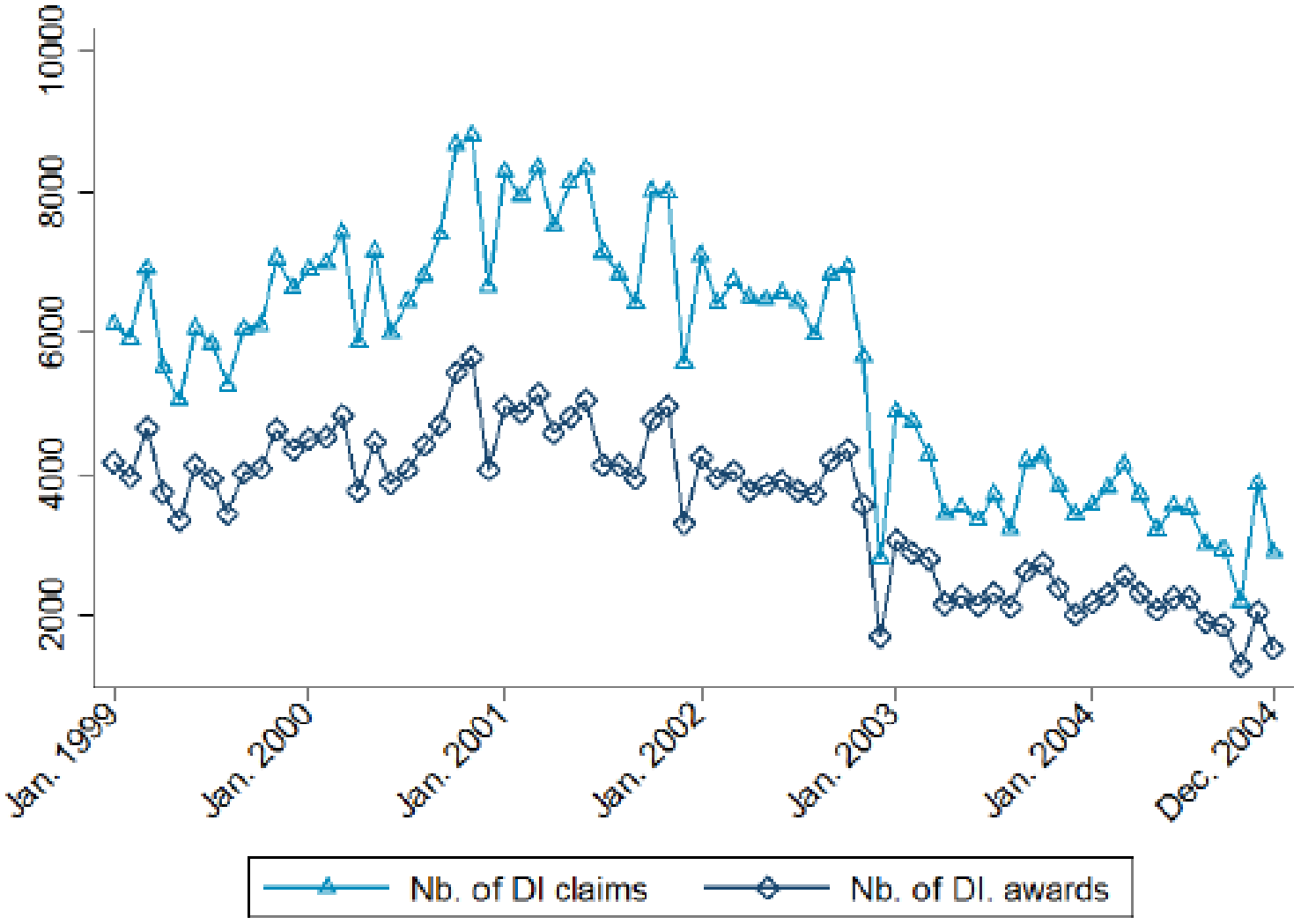
- So whether targeting improves is ultimately an empirical question

Data

- Linked administrative data from Statistics Netherlands:
  - Tax records register (earnings), from 1999- 2011
  - Municipality register (1995-2015)
  - Hospital discharge register (LMR), (1995-2005), ICD9 codes
  - Cause of Death Register (DO)
- Administrative data from National Social Insurance Institute (NSII), linked to Stat Neths databases.
  - Applications and award decisions (1999-2013)

Selection criteria

- Repeated cross section of individual in 2001-2004
  - Pre-gatekeeper: 2001-2002
  - Gatekeeper: 2003-2004
- Prime aged (25-64) and employed in previous year
- For each year we take all NSII application data and 10% random sample of non-applicants in Dutch population
- In total around 700.000 per year


2000
4000
6000
8000
10000
Jan. 1999
Jan. 2000
Jan. 2001
Jan. 2002
Jan. 2003
Jan. 2004
Dec. 2004
Nb. of DI claims
Nb. of DI. awards

Descriptive statistics

Table 1: Summary statistics of DI applicants and non-applicants on the month of (potential) application, before and after the introduction of the Gatekeeper protocol

	2001-2002		2003-2004		Diff-in-diff
	Applicants	Non-applicants	Applicants	Non-applicants	
	(1)	(2)	(3)	(4)	(5)
Demographics					
Age	43.30	38.37	43.81	39.02	-0.151***
	(9.82)	(9.46)	(9.88)	(9.53)	(0.044)
Male	0.41	0.56	0.48	0.56	0.075***
	(0.49)	(0.50)	(0.50)	(0.50)	(0.002)
Health status					
Health index	0.18	0.03	0.27	0.03	0.089***
	(0.88)	(0.34)	(1.14)	(0.34)	(0.002)
Dead within five years	0.03	0.01	0.04	0.01	0.011***
	(0.17)	(0.10)	(0.20)	(0.10)	(0.000)
Award decision					
Award rate (cond.	0.60	.	0.62	.	.
on applying)	(0.49)	.	(0.49)	.	.
Nb. of obs.	146,134	2,094,450	75,691	2,186,129	4,502,404

Empirical Strategy

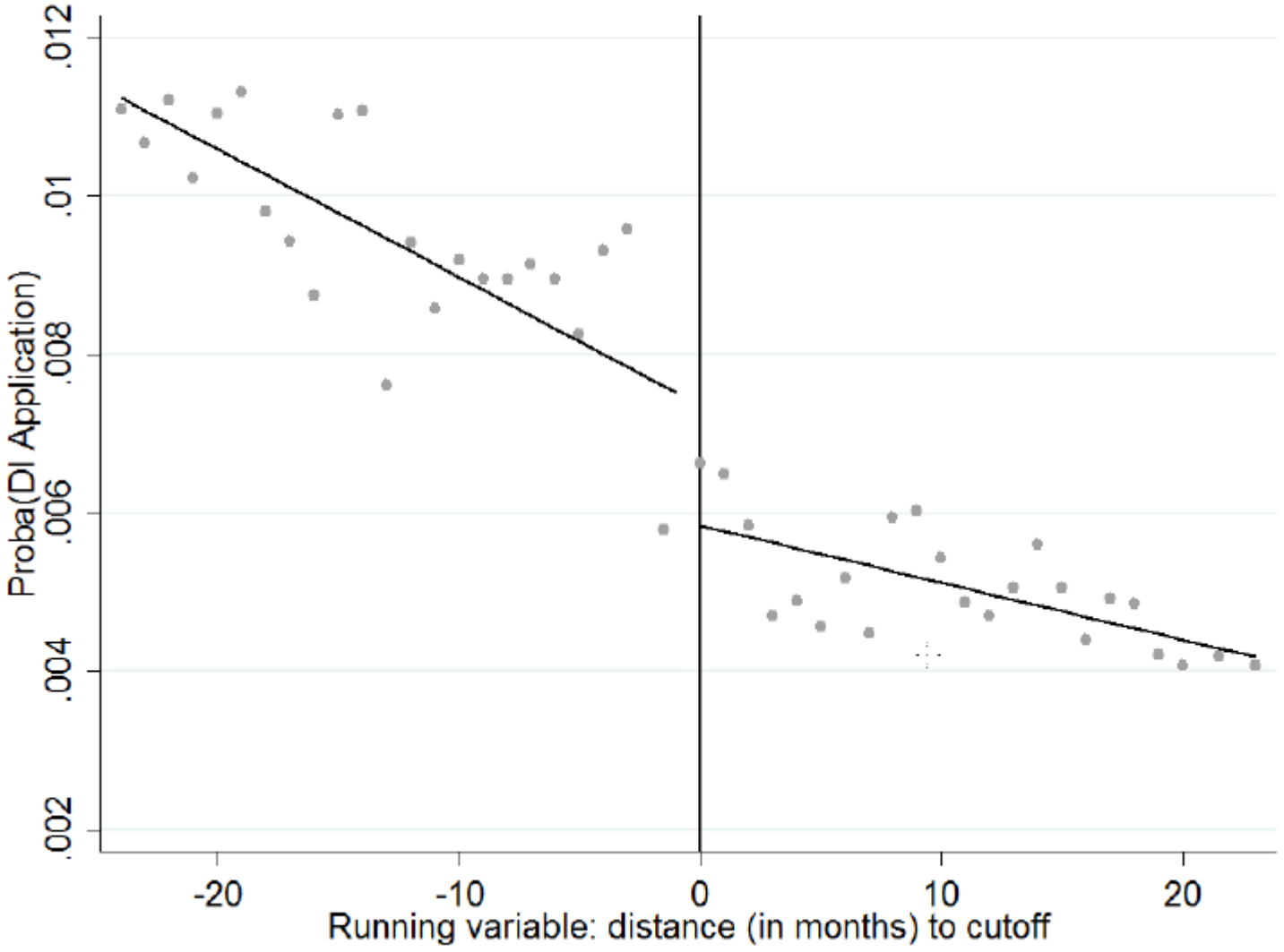
We employ two strategies

1. The causal effect of the introduction of the GKP
   - The extensive margin
   - What is the effect of the implementation of a system with structured reintegration process and gatekeeper screening?
2. Given a change in the system, what is the causal effect of marginal changes in screening intensity
   - The intensive margin
   - Insightful on underlying mechanisms
   - One region focused on employer, the other on the worker

The extensive margin: Regression Discontinuity in Time

- Intuition: use time as the running variable
- Assume that time varying confounders change smoothly across the date of the policy change
- Non-smooth changes across the threshold could be ascribed to the policy change
- We leave out nov-dec-jan: Donut hole regressions
  - To avoid anticipation effects (no evidence in raw data!)
  - Reorganization effects before gatekeeper came into effect
- We also exclude the two treatment regions (discuss later)

Figure 4: RD plot for DI application, linear fit



- More specifically, we estimate:

$$Y_{it} = c + \beta GKP_t + Time_t + \delta GKP_t * Time_t + Month + \varepsilon_{it}$$

- Where
  - *Y* is the outcome variable (application/award/share)
  - *GKP* dummie if t≥ 2003
  - *Time* a linear time trend , *Month* are month of year effects
  - Standard errors clustered at the individual level

- β is the parameter of interest (extensive margin treatment effect)
  - Is likely a work resumption effect (reintegration efforts) + decision to stop application procedure

- Donut to control for anticipation/noise around date of implementation
- Varying bandwidth/polynomial in time / placebo tests with alternative dates

Results extensive margin: change from old to new system

(what is the effect of the introduction of the new system?)

Table 3: RD and Donut-RD estimates. The effect of the Gatekeeper pr
on DI application rates.

	DEPENDENT VARIABLE **DI application**
	Coeff. [% change] (se)
RD estimate	-0.009* [-15.5%] (0.005)
Mean of dependent variable	0.058
Number of observations	2,237,250
Donut RD dropping those within one month of the cutoff	-0.024*** [-39.3%] (0.003)
Mean of dependent variable	0.061
Number of observations	2,257,510
Donut RD dropping those within two months of the cutoff	-0.027*** [-44.3%] (0.002)

Who stops applying

Table 4: The effect of the Gatekeeper protocol on DI application rates by impairment type. Donut-RD estimates.

	DEPENDENT VARIABLE **DI application**
	Coeff. [% change] (se)
Panel A: All impairments	-0.024*** [-39.3%] (0.003)
Panel B: Hard-to-verify impairments	-0.022*** [-41.3%] (0.002)
Musculo-skeletal	-0.008*** [-45.8%] (0.001)
Mental disorders	-0.008*** [-38.1%] (0.001)
"Other" disorders	-0.005*** [-38.5%] (0.001)
Panel C: Easy-to-verify impairments	-0.002*** [-32.6%] (0.000)
e.g. Cardiovascular diseases	-0.001*** [-35.8%] (0.000)
Nb. of obs.	2,071,474

$\Longrightarrow$ The sharp fall in applications primarily comes from individuals applying for hard-to-verify impairments (a -41.3% decrease).

Who stops applying?

- Females, prime aged and individuals who have better ex-ante health
- Biggest decline in hard to verify diseases

=> This generates changes in the composition of the pool of applicants

Can we characterize the leavers and the stayers?

(i.e. those also apply under GKP vs those who stop applying)

- We have:
  - Fall in DI application rate for a subgroup (say males)
  - Average share of males pre- and post GKP
  - Effect of reform on share of males

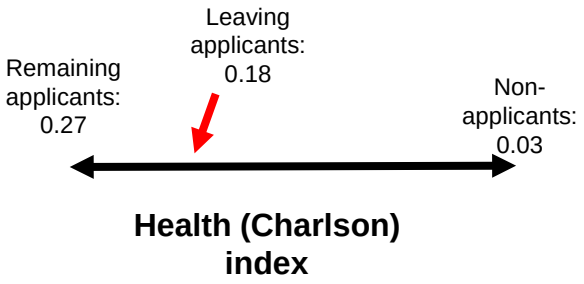
  With this information it is possible to solve for value avgs of *x* for those we keep on applying (x_{stay}) and those who stop applying (x_{leave}).

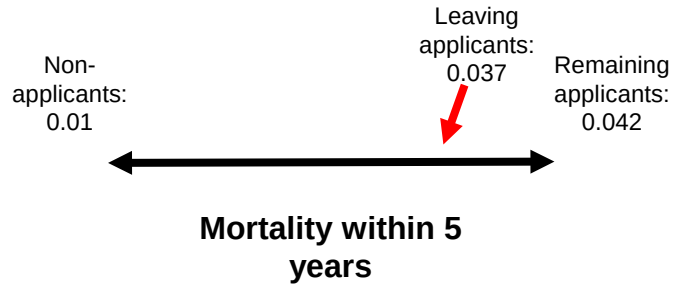
Composition of the applicant pool

Table 5: The effect of the Gatekeeper protocol on the composition of the pool of DI applicants – Donut-RD estimates.

	Change in average value/proportion of applicants [% change]	$x_{stayers}$	$x_{leavers}$	$x_s - x_l$
Panel A: Health and Mortality				
Health (Charlson) index	0.036*** [+18.6%]	0.27	0.18	0.087***
	(0.002)			(0.024)
Death rate within five years	0.002 [+6.0%]	0.042	0.037	0.004
	(0.002)			(0.005)

Less applicants in good health





Composition of pool of applicants: other characteristics

		x_{stay}	x_{leave}	
Panel B: Socio-Demographics				
Male	0.068*** [+16.5%]	0.48	0.32	0.166***
	(0.007)			(0.023)
Age				
Young (25-34)	0.027*** [+11.7%]	0.22	0.15	0.066***
	(0.004)			(0.013)
Prime age (35-49)	-0.026*** [-5.7%]	0.45	0.61	-0.063***
	(0.007)			(0.017)
Senior (50+)	-0.001 [-0.32%]	0.33	0.33	-0.003
	(0.005)			(0.013)
Past earnings (one-year lag)	657.75**	23,832.82	22,226.52	1606.30**
	(239.71)			(755.48)
Panel C: Share of impairment types				
Hard-to-verify impairments	-0.024*** [-2.7%]	0.86	0.92	-0.058***
	(0.008)			(0.008)
Panel D: Share awarded DI	0.062*** [+10.4%]	0.63	0.47	0.152***
	(0.007)			(0.010)
Nb. of individuals	107,297	38,886	29,525	68,411

- So, indeed changes in the pool of applicants
- Those who stop applying are on average healthier than those who remain
- Those who stop applying more often from difficult to verify conditions
- This does not rule out that some deserving individuals decide not to apply

  => look at non-applicants

- How do those who stop applying after GKP compare to those who did not apply in the prior to GKP?

Table 6: The effect of the Gatekeeper protocol on future outcomes (one-ye
lead) of **non-applicants** – Donut RD estimates.

Future (one-year lead):	Change in mean value/prop.	$x_{leavers}$	$x_{non-app}$	$x_l - x_{na}$
Panel A: Health and Mortality				
Hospitalization (any type)	0.004*** [+5.9%]	0.14	0.07	0.071***
	(0.001)			(0.)
Death rate	0.0002*** [+16.2%]	0.008	0.001	0.007***
	(0.0001)			(0.003)
Panel B: Labor-market outcomes				
Earnings	-437.7*** [-1.6%]	17,308	26,751	-9,442,26***
	(54.26)			(895.90)
Employment	-0.004*** [-0.4%]	0.83	0.92	-0.086***
	(0.001)			(0.023)
UI receipt	0.005^{μ} [+9.5%]	0.14	0.05	0.086***
	(0.003)			(0.021)
Welfare receipt	0.001* [+5.0%]	0.016	0.019	-0.003
	(0.000)			(0.006)
Nb. of individuals	2,150,213	29,525	1,061,504	2,150,213

Further gains in targeting efficiency at point of award decision?

- More information at the point of the award decision
- Does this lead to (further) improvement is targeting efficiency
- Can first see if there are additional composition changes:
  - See whether the avg health changes of awardees differ from avg health changes in applicants

Further gains at the point of the award decision?

Variable:	Pre-Gatekeeper			Post-Gatekeeper			
	$X_{app,0}$	$X_{award,0}$	δ_0	$X_{app,1}$	$X_{award,1}$	δ_1	$\frac{\delta_1}{\delta_0}$
Panel A: All impairments							
P(Health (Charlson) Index $\geq$ 1)	0.083	0.109	1,349	0.109	0.141	1.334	0.989
Dead within five years	0.033	0.045	1.39	0.042	0.056	1.38	0.992
Panel B: Hard-to-verify impairments							
P(Health (Charlson) Index $\geq$ 1)	0.058	0.075	1,32	0.079	0.102	1.32	1.003
Dead within five years	0.027	0.037	1.39	0.035	0.049	1.39	0.996
Panel C: Easy-to-verify impairments							
P(Health (Charlson) Index $\geq$ 1)	0.269	0.290	1,11	0.297	0.322	1.13	1.016
Dead within five years	0.074	0.085	1.17	0.080	0.094	1.20	1.028

Reading note: $X_{app,0}$ stands for the average value of X for applicants in the pre-Gatekeeper period.

- No further gains in Targeting efficiency
  - Changes in the pool of awardees are completely driven by self-screening (incl work resumption during waiting period)

Results from a field experiment: the intensive margin

(what, given the system, is the effect of intensifying screening?
Are there further improvement in targeting efficiency?
What can be learned from this (mechanisms)?)

More on the experiment

- Experiment started on January 2003, when first applications arrived and ended in October 2003

  => New to all involved (worker/employer/doctor)
- We *exogenously* change screening intensity in two regions:
  - In 24 out of 26 regions standard approach: test on paper
  - Caseworkers in remaining 2 regions instructed to always implement a stricter screening of applications: contact employer/worker (unless obvious case)

  - Avg time spend on screening is 40% higher in experiment regions
  - In treatment regions more direct contacts with employer and worker

Checking procedures compared

Question: is screening stricter in experiment regions? Answer: yes)

Table 1: Difference in screening stringency between treatment and control regions

	Treatment regions		Control
	Apeldoorn	Hengelo	regions
Only on paper	4%	14%	25%
Telephonic contact with employer	33%	34%	52%
Telephonic contact with worker	14%	14%	23%
Telephonic contact with occupational health agency	3%	12%	32%
Visit to employer	9%	41%	7%
Face-to-face contact with worker	77%	41%	7%
Unknown	4%	2%	

Note that caseworkers can use multiple screening methods on one application, so columns can add up to more than 100%.

The intensive margin: Difference-in-Difference

$$Y_{irt} = \alpha_1 + \alpha_2 Treated_{rt} + \delta_t + \delta_r + \alpha_3 X_{irt} + \varepsilon_{it}$$

- Y is the outcome variable (Apply/Award/Share)
- Treated is an indicator for being in the treatment area in 2003 (α_2 the intensive margin effect)
- δ_t year effacts (extensive margin!), δ_r region specific effects
- X a vector of individual controls (Gender, age, ethnicity)
- Standard errors clustered at the regional level

- Extended DiD for test on common trend assumption (OK!)
- Allow for different treatment effect by region and year

Table 12: At the intensive margin: The effect of stricter screening of reintegration efforts (by NS caseworkers) on DI application rates – DiD estimates for 2003.

	DEPENDENT VARIABLE: DI application
	Coeff. [% change]
For impairment type:	(se)
All	-0.003* [-4.5%]
	(0.002)
Musculo-skeletal	-0.0005 [-2.6%]
	(0.001)
Mental disorders	-0.004** [-16.7%]
	(0.003)
Cardiovascular diseases	-0.000 [-3.3%]
	(0.000)
Nervous disorders	-0.000 [-0.1%]
	(0.000)
Respiratory disorders	-0.000 [-0.2%]
	(0.000)
Endocrine problems	-0.000 [-17.2%]
	(0.000)
Other	0.001 [+7.3%]
	(0.001)
Month*year dummies	✓
Regional dummies	✓
Background individual characteristics	✓
Nb. of obs.	3,563,624

Conclusions

- The introduction of the reform was very effective (i.e. sharp reduction in applications)
- This was accompanied by changes in the composition of the pool of applicant
- In accordance with expectations:
  - Pool of applicants becomes more deserving
  - Applicants are less healthy and a drop in conditions that are difficult to verify (mental, musculo-skeletal)

- Seen from this perspective the reform has improved targeting efficiency

- This self-screening effect can be interpreted as effect of increased work-resumption + decision of worker/employer to pull out of application process
- However, while those who stop applying are on average healthier than those who keep on applying, we also see that compared to non-applicants
  - Those who stop applying have worse health and future mortality
  - Lower income, lower employment rates, more often unemployed and on welfare
  - This might hint at workers pulling out, without work resumption

- The award decision at the end of the process did not lead to additional changes in the composition of workers
- This suggests that the GKP protocol of reintegration efforts in the waiting period is most effective in improving targeting efficiency
  - And not so much the increase in information accessible to the medical examiner at the point of the award decision)
- Further intensifying screening helps to additionally reduce applications, in particular for mental diseases
  - That are hard to verify and where it is more difficult to assess the severity or the appropriate measures