MONASH University

MONASH BUSINESS SCHOOL

CENTRE FOR HEALTH ECONOMICS

WORKING PAPERS

Does Stricter Disability Screening Lead to Improved Targeting on Long-term Health and Employment Outcomes?

Centre for Health Economics, Monash University
no. 2026-12

Chiara Campana, Pierre Koning, Maarten Lindeboom

Keywords: Disability Insurance, (Mental) health, Screening, Targeting

JEL Classification: H2, I3

Does Stricter Disability Screening Lead to Improved Targeting on Long-term Health and Employment Outcomes?*

Chiara Campana† Pierre Koning‡ Maarten Lindeboom§

August 27, 2026

Abstract

While evidence points at potentially strong effects of stricter screening and tightening eligibility criteria for Disability Insurance (DI), little is known about their targeting effects: are screened-out workers also those with better future health and more ability to work? To shed light on this, we employ Regression-Discontinuity-in-Time regressions that compare the long-term outcomes of DI application cohorts just before and after a reform in the Netherlands. The reform led to stricter screening in the sickness period before DI application and reduced the number of applicants by 33%. Up to 18 years after application, we find persistently lower survival rates and lower employment rates of post-reform cohorts. Using detailed information on the future chronic diseases of these cohorts and the application diagnoses, we next construct indices capturing long-term expected survival (a "Survival Health Index", SHI) and long-term employment prospects (a "Work Ability Index", WAI) for DI applicant cohorts in the years following application. These indices allow us to assess whether the reform primarily deterred applicants with health-related disabilities, work-related disabilities, or both. We find that the self-screening induced by the reform was stronger for applicants with work disabilities (low WAI scores compared to their SHI scores) than for health disabilities (low SHI scores compared to their WAI scores). This is largely due to a decline in applications of workers with mental and musculoskeletal health conditions that have high survival rates, but lower employment rates.

JEL code : H2, I3

Keywords : Disability Insurance, (Mental) health, Screening, Targeting

---

*We thank participants from the Ageing and Health NBER workshop in Lugano in 2025, the NBER Summer Institute workshop of 2025 on Social Security in Boston, the AHEPE/IRDES workshop in Paris of 2025, the Essen Health Conference in 2025, the NETSPAR Pension Day in Utrecht in 2024, the 2025 Summer Institute on Ageing at Venice International University, the EuHEA Online seminar series (in spring 2025), the online IPRA doctoral tutorial of 2025, the LOLA-HESG 2025 conference in Ermelo, and the lunch seminar series of the department of economics of the VU Amsterdam, Ca'Foscari University in Venice and Tinbergen Institute for helpful comments to the paper. Special thanks go to Anneke van der Giezen, Katja Hanewald, Martin Karlsson, Marloes Lammers, Gijsbert van Lomwel, Fabrizio Mazzona, Roger Prudon, Sannah van der Wal and Nicolas Ziebarth for comments and suggestions to earlier drafts. Disclosure statement: The authors herewith state to have no relevant or material financial interests that relate to research described in this paper. The authors use confidential administrative data owned by Statistics Netherlands. The data can be accessed only by remote access connection, after filing a request directly to Statistics Netherlands (https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/aanvraag-toegang-microdata). The authors are willing to assist in doing so. An IRB approval was not mandatory for this research project, as all privacy conditions and ethical standards are warranted by the owners of the data, Statistics Netherlands.

†Vrije Universiteit Amsterdam, Ca' Foscari University of Venice. Email : c.campana@vu.nl

‡Vrije Universiteit, Amsterdam, IZA@LISER and TI

§Vrije Universiteit, Amsterdam, TI, Centre for Health Economics, Monash University, IZA@LISER.

1 Introduction

Disability Insurance (DI) enrollment for workers has been on the rise in many OECD countries (OECD, 2022). This trend is difficult to explain from worsening health conditions of individuals alone and raises questions about the targeting of DI programs: should DI benefit provide income to workers with severe health conditions only ("health disability"), or to those with insufficient earnings capacity ("work disability") (Benítez-Silva et al., 2014; Godard et al., 2024; Haller et al., 2024; Liebert, 2019)? Knowing that DI receipt is mostly permanent for workers but health conditions may change, it is key to understand the long-term consequences of DI reforms on the composition of cohorts of DI applicants and awardees: how do these affect average health, mortality and employment outcomes? This is in essence what this paper is about.

In this paper, we use a Dutch reform that tightened the screening of DI applications to study the long-term evolution—up to 18 years after application—of chronic conditions, mortality, and employment. In January 2003, the "Gatekeeper Protocol" (GKP) reform substantially raised application costs for both workers and employers by intensifying medical checks and work-resumption efforts during the one-year sickness/waiting period preceding a DI application. As a result, the reform discouraged a large group of workers from applying for DI benefits. Building on Godard et al. (2024), we first infer the long-run survival and employment trajectories of applicants who are screened out by the reform (compliers) and compare them to those who continue to apply after the reform (non-compliers). Using rich information on chronic conditions and medical diagnoses at the time of DI application, we then construct indices capturing long-term expected survival (a "Survival Health Index", SHI) and long-term employment prospects (a "Work Ability Index", WAI) for DI applicant cohorts in the years following application. These indices allow us to assess whether the reform primarily deterred applicants with health-related disabilities, work-related disabilities, or both.

Using Regression-Discontinuity-in-Time (RDiT) specifications around the GKP reform cutoff, we first document a sizeable 33% decline in the DI application rate. This drop is accompanied by lower survival rates up to 18 years after the DI award decision, amounting to a cumulative 1.1 percentage point reduction (about 12% of the baseline mortality rate). Over the same horizon, we also find that employment rates for the post-GKP cohort are around four percentage points lower (roughly 6% of the baseline employment rate). Next, we employ Almond and Doyle (2011) to infer the characteristics of compliers to the reform, i.e. those who no longer applied for DI. In the first year after application, the implied survival rates of these compliers are similar to those of employed workers who never applied for DI, indicating effective short-run targeting on health. From the second year onward, however, the inferred survival and employment trajectories of compliers resemble those of workers who continued to apply after the reform (non-compliers) but were denied benefits. This indicates that long-term targeting improvements were limited.

To more precisely capture compositional changes that may underlie survival and employment outcomes, we next construct the Survival Health Index (SHI) and the Work Ability Index (WAI) as summary measures of the Health and Work limitations of DI applicant cohorts. To do so, we use rich information on post-decision medical prescriptions from a national medical prescription database and link these to DI applicants in the years following their award decision. We map the prescriptions into a standardized set of 22 chronic disease categories and combine this information with physician diagnoses recorded at application by the assessors of the Dutch National Social Insurance Institute (NSII) to construct our indices.[1] For the construction of the SHI, we follow Danesh et al. (2024) to estimate 10-year survival rates as a function of chronic diseases and NSII-assessor diagnoses with LASSO for the subsample of individuals observed at age 65.[2] The estimated coefficients from this model are then used to predict expected survival rates beyond age 65 for all, mostly younger, applicants in our sample. This procedure enables us to detect variation in life expectancy for all applicants that would otherwise go unnoticed.[3] For the WAI scores, we estimate annual employment outcomes up to 18 years after the DI award decision using the same information on chronic conditions and application diagnoses.[4]

Combining information from the SHI and WAI, we infer that applicants with mental and musculo-skeletal disorders are more likely to exhibit work disabilities (low WAI) than severe health impairments that limit life expectancy (low SHI). More strikingly, applicants with low WAI scores and high SHI scores (i.e. high survival but low employment rates) display substantially more comorbidities than applicants with high WAI scores and low SHI scores. This pattern suggests that comorbidities can have a large combined impact on work capacity, while their joint adverse effect on life expectancy remains relatively limited. Consistent with the fact that the

[1] In Dutch "UWV". UWV This is the Institute that is responsible for the admission process and the administration of DI Benefits. See also Section 2 for more detail of the DI award process.

[2] Danesh et al. (2024), use the same chronic disease data to construct a "Survival Health Index" (SHI) as a proxy of health gradients.

[3] Mortality rates of younger workers are too small to obtain meaningful estimates.

[4] The idea of constructing distinct indices for disability is similar to Croda et al. (2018), who derive indices for overall health, work capacity, functional limitations, and mental health. Their approach uses principal component analyses to obtain indices, rather than regressions on long-term survival and employment outcomes.

GKP reform disproportionately screened out workers with mental conditions, we conclude that (self-)selection operated more strongly on work disabilities than on underlying health disabilities that reduce life expectancy.

This paper adds to a growing body of research that examines the effects of stricter screening for DI applications or tightening eligibility criteria for DI benefits (Ahammer and Packham, 2025; De Jong et al., 2014; Haller et al., 2024; Johansson et al., 2014), as well as the effects of changes in application costs (Deshpande and Li, 2019; Godard et al., 2024).[5] A key finding in this literature is that stricter eligibility criteria can significantly reduce the inflow into DI schemes along two channels: (i) the direct impact of stricter screening and (ii) the impact of self-screening, where workers are discouraged to apply. Haller et al. (2024), who study the effect of increased screening for DI in Austria, argue that a notable proportion of compliers to the reform return to the labor market, indicating improved targeting and minimal fiscal spillovers to other welfare programs. Although these studies focus primarily on short-term employment, a comprehensive evaluation of the targeting effects requires insight into long-term health and employment outcomes following the DI award decision. Our study addresses this gap in the literature.

Our paper also contributes to the recent literature on the causal effects of DI benefit levels and receipt on mortality (Black et al., 2024; Garcia-Gomez and Gielen, 2017; Gelber et al., 2023). DI programs compensate for income loss associated with adverse health conditions, potentially preventing further health deterioration. However, DI benefit receipt may also weaken work incentives, potentially affecting employment, health, and mortality outcomes. Using exogenous variation in the assignment of medical examiners for SSDI applicants, Black et al. (2024) find that mortality increases for applicants at the margin of being allowed benefits but decreases for infra-marginal applicants who receive benefits. Gelber et al. (2023) demonstrate that increases in SSDI benefits lead to lower mortality, with an estimated elasticity of mortality with respect to DI income ranging from 0.6 to 1.0. It should be noted, however, that mortality is a coarse measure of health, particularly for mental health conditions, which account for roughly one-third of DI inflow in most OECD countries (OECD, 2010). This underscores the need to examine health outcomes more comprehensively to better understand the full impact of DI programs—which is what we do in the current paper.

[5] See Godard et al. (2024) for an extensive review.

2 The Gatekeeper Protocol

Workers in the Netherlands are subject to a waiting period before they can apply for DI benefits.[6] During this period, permanent workers continue to receive full wage payments, while temporary workers are eligible for Sickness Insurance (SI) benefits that amount to 70% of their wage earnings from the moment their contract ends. At the time of the Gatekeeper Protocol (GKP) reform in 2003, the sickness waiting period was 12 months.[7] At the 12-month assessment, a medical examiner from the Employee Insurance Agency (UWV) evaluated the worker's impairments, the extent to which these conditions restrict the type and amount of work that the claimant can perform and the (expected) permanence of these limitations. As will be explained later on, we proxy the longer term health and work ability of claimants with our Survival Health Index (SHI) and a Work Ability Index (WAI). The assessment determined the worker's remaining earnings capacity, defined as the gap between the worker's last earned wage and the average wage of the five highest-paying jobs the individual is still deemed able to do. Individuals who had lost at least 80% of their earnings capacity were classified as fully disabled and received DI benefits equal to 70% of their pre-disability earnings. Partial DI benefits amounted to 70% of lost earnings capacity for disability degrees between 15% and 80% of pre-disability earnings, while applicants with a disability degree below 15% were not eligible for DI benefits.

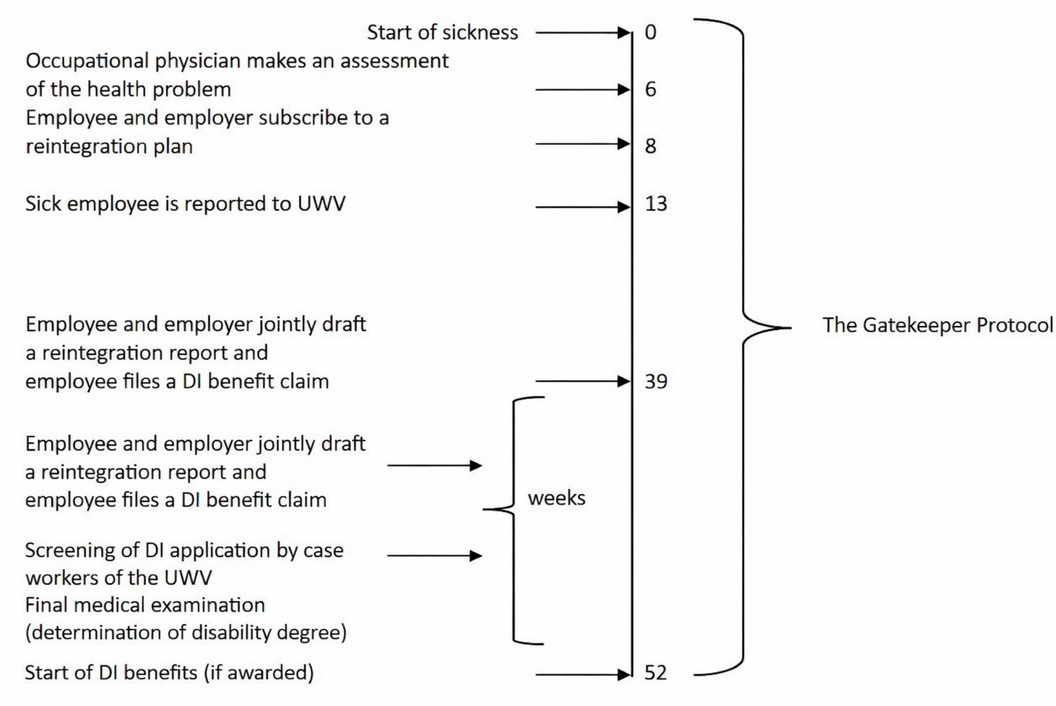
The Gatekeeper Protocol (GKP) reform introduced in 2003 is widely regarded as the most effective policy in the Netherlands for reducing disability insurance (DI) inflow and enrollment (Koning and Lindeboom, 2015). The reform applied to workers whose sickness absence began in January 2002 or later and whose DI assessment occurred from January 2003 onward. Although the reform did not modify formal DI eligibility criteria, it substantially increased the responsibilities and financial incentives for both employers and employees in the waiting period. At the time, DI enrollment had reached 12% of the insured population, raising concerns about insufficient screening of health conditions during the waiting period. In response, the GKP introduced a structured set of mandatory reintegration steps that employers and employees must complete before a DI application can be filed after 12 months of absence. Figure 1 illustrates the GKP process timeline, highlighting its key milestones: contacting an occupational doctor after six weeks of sickness, drafting a reintegration plan after eight weeks, timely reporting of

[6] This section describes the main elements of the Dutch institutional settings for sickness, DI applications and DI benefits at the GKP reform. For more detailed information and related reforms, we refer to Koning and Lindeboom (2015), Godard et al. (2024) or Koning and Vethaak (2021).

[7] In 2004 the waiting period was extended to two years (Koning and Lindeboom, 2015). In this section, we describe the sickness and disability system as it was at the time of the GKP reform.

the absence to the NSSI after 13 weeks, and submitting a completed report after 39 weeks[8]

Figure 1: Timeline of one-year sickness period before DI applications



By strengthening monitoring and accountability, the reform aimed to intensify reintegration efforts. To the extent that these efforts are effective, stricter screening should increase return-to-work rates during absence and reduce subsequent DI applications. This mechanism represents the direct effect of the reform. Increased reintegration requirements also increase the costs of applying for disability insurance (DI) and reduce the noise in the disability signal. As a result, some workers with sufficient work capacity may opt to resume work rather than continue the application process, reflecting a *self-screening* mechanism. This (second) mechanism is likely to be strong for conditions that are relatively easy to verify, where diagnosis, treatment, and short-term prospects for resuming work are relatively clear. Such conditions may include acute and temporary health problems with predictable recovery paths. On the other hand, easily verifiable impairments typically also encompass severe conditions—such as advanced-stage cancers or serious cardiovascular diseases—that effectively preclude a return to work. In these cases, the marginal impact of stricter screening is mostly limited. Overall, the reform may therefore improve the targeting of DI benefits.

[8]This system has remained largely unchanged since its introduction in 2003.

Given the substantial additional effort required to comply with the obligations during the one-year sickness period, compliers with the GKP reform can be understood as workers who either return to work during the waiting period or perceive the requirements as too burdensome relative to their health condition. In the former case, the reform likely improves the targeting of DI benefits by encouraging work resumption among those with remaining work capacity. Using data on hospitalizations prior to DI application, Godard et al. (2024) provide evidence consistent with this mechanism. They document a sharp decline in DI applications—nearly 40% among permanent workers—with the largest reductions occurring among individuals with hard-to-verify conditions, such as mental health disorders, burnout, depression, and musculoskeletal problems. For these conditions, stricter protocols reduce informational frictions in the assessment process and improve the accuracy of award decisions. However, higher application costs may also discourage "truly" disabled workers from applying. This concern is particularly relevant for individuals with limited expected remaining lifespan, for whom the returns to completing the application process are relatively low and costs are high. In such cases, the reform may induce disproportionate exits among the truly disabled compared to the able, described by Parsons (1991) as perverse self-screening.

From a health perspective, the targeting properties of the program following the GKP reform can be assessed by examining the long-term health and survival outcomes of compliers. The onset or progression of health conditions may reduce well-being and shorten expected residual life years—both in good and poor health—among non-compliers (i.e., "stayers" who continue to apply after the reform). For some conditions, such as Burnout, Depression, Musculo-skeletal diseases etc, the severity of the condition may be hard to verify. If such "hard-to-verify" conditions are associated with lower mortality than more easily verifiable diseases, self-screening under the GKP may have improved targeting in terms of these health outcomes. However, evaluating targeting through an employment lens—centered on the ability to work—is also important. From this perspective, the implications of self-screening are less clear and depend critically on how different conditions affect employment capacity and how permanent this condition is. This motivates not only an analysis of long-term survival and employment outcomes among DI applicants and beneficiaries, but also an examination of the underlying health trajectories that shape these outcomes, as captured by the concepts of "work ability" (the Work Ability Index, WAI) and "health disability" (the Survival Health Index, SHI), which will be discussed in section 5.

3 Data

3.1 Data sources

For our empirical analysis, we use data from different administrative sources provided by Statistics Netherlands (CBS): the Municipal Personal Records Database, employment administration register, the Medical Prescription database and data from the Netherlands Employee Insurance Agency (UWV). These data sources are merged through a unique identifier.[9] The UWV database contains information on all DI applications and awards, including seven broad diagnosis categories reported at the time of application: musculo-skeletal, mental, cardiovascular, nervous, respiratory, endocrine, and other non-categorized disorders. It also records the assessed degree of disability for recipients of partial DI benefits (15–25%, 25–35%, 35–45%, 45–55%, 55–65%, and 65–80%) and full benefits (80–100%). We observe the exact date of the award decision, but not the start date of the sickness spell. As a result, the lag between the application date, which occurs in week 39, and the award decision—made at most two months later—is not directly observed. In our empirical analysis, we therefore follow Godard et al. (2024) and estimate a Regression-Discontinuity-in-Time (RDiT) specification with a one-month donut hole on either side of January 2003, excluding applicants whose award date cannot be assigned with sufficient certainty to either the pre- or post-GKP period.

From the Municipal Personal Records Database of Statistics Netherlands, we obtain information on gender, date of birth, nationality, marital status, household composition, place of residence for all individuals living in the Netherlands, and, where relevant, dates of death. Annual earnings from the employment administration register of Statistics Netherlands are observed from 1999 to 2022, together with individuals' yearly economic status: employee, self-employed, unemployment insurance, social assistance, pension, DI, student, or other without income.

Finally, we observe annual prescription data from 2006 to 2022, which record reimbursements for medications covered under the statutory basic health insurance program. The prescriptions are classified into ATC codes, the Anatomical, Therapeutic, and Chemical classification system of the World Health Organization. Following Huber et al. (2013), we map these prescription records to 22 chronic disease conditions, as shown in Appendix Table B1.

[9] Statistics Netherlands (CBS) databases assigns a random identification number from the citizens' social service number (BSN).

3.2 Sample selection

We merge and select our data from different sources in a number of steps. First, the sample of DI applicant cohorts we follow is restricted to the period 2001-2004 to examine two years before and after the reform. This yields a sample of 449,017 DI applicants. We exclude workers who applied for DI benefits more than once or whose records contain conflicting information. We also exclude applicants with multiple records for the same award year and award date, as well as those with missing diagnoses or award decisions. These exclusions reduce the sample to 355,479 individuals. We then further restrict the sample to applicants aged 25–65 who were formally employed and not receiving DI benefits in the year prior to the award decision, which reduces the sample to 290,813 applicants. An applicant is classified as employed if he/she worked for at least six months in a given year; accordingly, the sample may also include part-time and temporary workers.

Table 1 presents descriptive statistics for DI applicants two years before and after the GKP reform. Columns (1) and (2) indicate that the reform coincided with shifts in applicant characteristics, with the difference in means reported in column (3). After the GKP reform, DI applicants tend to be older, are more likely to be male and non-native, have lower annual earnings, and are less likely to be employed in the year of the award decision. In contrast, in the year prior to application, post-GKP applicants display slightly higher earnings than their pre-GKP counterparts. With respect to application diagnoses, we observe a sizable decline in the share of applicants with mental disorders after the reform, falling from 37% to 28%. The DI award rate also increases after the reform, rising from 61% to 64%. This increase is largely driven by awards of full DI benefits, suggesting a rising share of applicants with no or limited remaining earnings capacity.

Table 1: Summary statistics of pre- and post-GKP reform DI applicant cohorts (2001-2004)[(a)]

	2001-2002	2003-2004	Difference
	Pre-GKP applicants	Post-GKP applicants	in-means[(b)]
	(1)	(2)	(3)
Demographics			
Age	43.07	43.70	0.635***
	(9.65)	(9.71)	(0.008)
Male	0.402	0.473	0.072***
	(0.490)	(0.499)	(0.001)
Native	0.805	0.783	-0.022***
	(0.396)	(0.412)	(0.001)
Labour market outcomes[(b)]			
Employed (year award decision)	0.820	0.761	-0.068***
	(0.385)	(0.427)	(0.002)
Gross earnings (2015 euros)			
— in year of award decision	25,841	24,364	-1,479.63***
	(23.312)	(20.036)	(87.42)
— in year before award decision	30,238	30,847	609.20***
	(17.389)	(18.021)	(69.20)
Application diagnoses			
– Cardiovascular	0.043	0.056	0.013***
	(0.206)	(0.232)	(0.001)
– Diseases of pregnancy	0.016	0.080	-0.009***
	(0.126)	(0.089)	(0.001)
– Diseases of the blood	0.004	0.007	0.003***
	(0.067)	(0.084)	(0.001)
– Diseases of the digestive system	0.017	0.021	0.005***
	(0.130)	(0.147)	(0.001)
– Diseases of the ear	0.003	0.003	0.001***
	(0.058)	(0.060)	(0.001)
– Diseases of the eye	0.004	0.005	0.002***
	(0.065)	(0.075)	(0.001)
– Diseases of the genitourinary system	0.022	0.026	0.005***
	(0.147)	(0.161)	(0.001)
– Diseases of the nervous system	0.035	0.044	0.009***
	(0.183)	(0.206)	(0.001)
– Diseases of the respiratory system	0.013	0.015	0.001***
	(0.113)	(0.122)	(0.001)
– Diseases of the skin	0.005	0.005	0.001***
	(0.069)	(0.069)	(0.001)
– Endocrine disorders	0.011	0.013	0.003***
	(0.104)	(0.114)	(0.001)
– Mental	0.370	0.291	-0.087***
	(0.483)	(0.451)	(0.001)
– Musculoskeletal	0.294	0.292	-0.003***
	(0.456)	(0.455)	(0.001)
– Others	0.159	0.217	0.058***
	(0.366)	(0.412)	(0.001)
DI awards	0.623	0.642	0.019***
	(0.485)	(0.479)	(0.002)
– Partial DI benefit	0.308	0.304	-0.004**
	(0.462)	(0.460)	(0.002)
– Full DI benefit	0.307	0.334	0.027***
	(0.461)	(0.472)	(0.002)
Total number of observations	191,570	99,243	290,813

Notes: [(a)] Our sample considers yearly observations and excludes applicants not employed in the previous year; [(b)]: Column (3) shows the mean differences and their standard errors.

3.3 Long-term outcomes: survival, employment and chronic diseases

In section 5 we will define use two indices that are constructed from the survival rates and employment outcomes of pre- and post-reform applicant cohorts: for health the Survival Health Index (SHI) and for work the Work Ability Index (WAI). Appendix Figure A1 illustrates the evolution of survival and employment rates up to 18 years after application for the pre- and post GKP cohorts. The left-hand figure displays a widening gap in survival rates of about three percentage points over time. Although this suggests that compliers to the reform had substantially higher survival prospects than those who stayed in the system, the gap may also reflect underlying trends unrelated to the GKP reform. To further examine potential drivers of this cohort difference, Appendix Figure A2 presents survival estimates stratified by gender, age cohort, degree of disability, and application diagnoses. Conditioning on application diagnoses attenuates the survival gap between pre-GKP and post-GKP cohorts the most. This is consistent with the idea that the reform induced greater self-screening along dimensions of health and work capacity at the time of application.

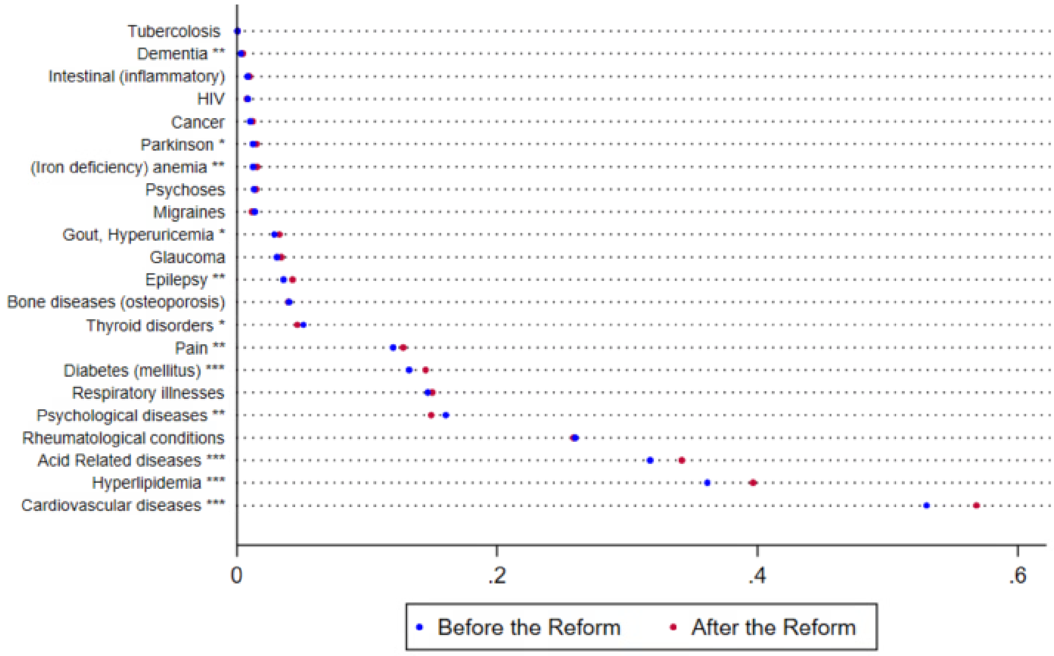
To gauge changes in work ability between pre- and post-reform cohorts, Appendix Figure A1b plots employment rates by elapsed time since application. We restrict the sample for each elapsed year to applicants who are alive and at most 65 years of age, the official retirement age for the cohorts under study.[10] For both the pre- and post-GKP cohorts, employment rates drop most sharply within the first four years after application, by about 20 percentage points. For all subsequent years, post-GKP cohorts consistently exhibit employment rates that are four to five percentage points lower than those of pre-GKP cohorts.

While survival rates offer a useful composite measure of overall health, they do not reveal changes in specific underlying conditions that may have resulted from the GKP reform. To explore such condition-specific effects, we analyze medical prescription data. Following Huber et al. (2013), we classify prescriptions into 22 chronic conditions (see Appendix Table B1 for details). Figure 2 presents the prevalence rates of these chronic diseases at age 65, ordered from least to most common, for the pre- and post-GKP reform applicant samples. It also reports the outcomes of equality tests comparing these rates across the two periods.

Figure 2 indicates that the GKP reform led to substantial and statistically significant shifts in the prevalence of several common chronic conditions—most notably acid-related disorders,

[10] After age 65, entitlement to DI benefits ends and individuals are in principle eligible for old-age pensions.

Figure 2: Prevalence (fractions) of chronic diseases at the age of 65 in DI application sample of pre-GKP (2001-2002) and post-GKP (2003-2004) cohorts



Notes: Prevalence rates of chronic conditions are derived for the sample of workers in pre- and post-GKP cohorts who turn 65 between 2006 and 2012. Asterisks indicate the statistical significance of a two-sample t-test for the difference in prevalence before and after the reform. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

hyperlipidemia, and cardiovascular diseases. An exception is psychological disorders, proxied by the use of antidepressants, for which no significant change in prevalence is observed.

4 GKP reform effects: Survival and employment

The GKP reform was implemented in April 2002 and began affecting new DI applicant cohorts from January 2003 onwards. Our empirical strategy exploits the resulting sharp discontinuity in January 2003 by applying a Regression Discontinuity-in-Time (RDiT) design (Hausman and Rapson, 2018).[11] Using this approach, we examine the effects of the GKP reform on application rates and on the composition of applicant cohorts.

Our analysis proceeds in three steps. First, to assess the share of *compliers*—individuals who stopped applying for DI benefits after the reform— we compare DI application rates before and after January 2003. Second, we estimate changes in the composition of DI applicants induced

[11] Alternatively, one could adopt a difference-in-differences (DiD) approach, comparing application cohorts immediately before and after the reform. We prefer the RDiT framework because it does not rely on the common trend assumption and yields more precise estimates of application responses and compositional changes.

by the GKP reform by analyzing survival and employment outcomes of pre- and post-reform cohorts up to 18 years after application. Finally, we use these estimates to characterize the compliers and compare their long-term survival and employment outcomes to those of *stayers* (who continued to apply) and *non-applicants*. This comparison sheds light on the longer-term targeting effects of the GKP reform.

4.1 The RDiT Model

To estimate the effect of the GKP reform on DI application rates, we analyze monthly data on *all* employed workers from 2001 to 2004, providing a symmetric 24-month window around the January 2003 reform cutoff. We evaluate the reform's effects on *survival* and *employment outcomes* up to 18 years after application by applying the same RDiT specification, but restricting the sample to *application* cohorts. Because we observe only the date of the DI award decision—not the exact application date—some uncertainty arises for decisions around the reform threshold. In particular, it is unclear whether sickness spells for individuals with award decisions in December 2002 began before April 2002 (and thus fall under the old regime). Following Godard et al. (2024), we therefore exclude observations from December 2002 and January 2003.

Our preferred specification is a linear RDiT with a bandwidth of 24 months on each side of the cutoff and a one-month "donut hole" around the threshold.[12] The linear probability model for the *monthly DI application rate* of all employed workers and the *survival* and *employment* outcomes among DI applicants is specified as follows:

$$Y_{it} = \alpha + \beta_Y \ T_i + \gamma_1 \ t + \gamma_2 t * T_i + \sum_{k=2}^{12} \delta_k M_{kt} + \epsilon_{it}, \tag{1}$$

where Y_{it} is the outcome variable that can be an indicator for a DI application of worker i $(i = 1, ..N)$ in the month of the year t. T_i is a dummy variable that takes the value of one after December 31, 2002, and it is equal to zero before the reform. We allow the linear time trend to differ on each side of the cutoff by including t and the interaction term $t * T_t$, where t equals zero at the cutoff and ranges from -24 to 24. The parameter γ_1 captures any pre-existing linear trend in DI applications, and γ_2 captures any differences with the linear trend after the policy change. M_{kt} with $k = 2, \ldots, 12$ represents a set of calendar-month indicators and δ_k month-of-year fixed

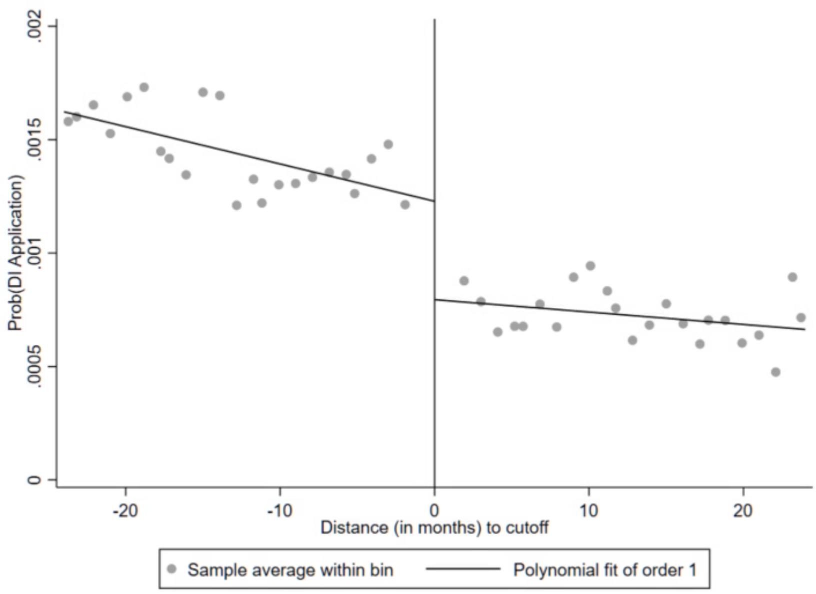
[12] As shown later, estimates are stable across bandwidths between 12 and 24 months; see Appendix Figure A3. Results for alternative specifications with varying bandwidths and higher-order polynomials appear in Figures A4 and A5 in Appendix A.

effects. Month-of-year fixed effects are included to control for any seasonal and administrative effects that may be relevant at the end of the calendar year. Standard errors are clustered at the month-year level. Analogous models are applied for employment and survival outcomes up to 18 years after application. So, the parameter of interest β_Y reflects the overall impact of the reform on the probability to apply for a DI benefit or survival and employment outcomes.

4.2 The GKP effect on DI applications

Figure 3 below displays the results for applications from our preferred donut RDiT regression, using monthly DI applications data around the reform cutoff date of January 2003. The monthly rate is calculated as the fraction of applicants divided by the total employed Dutch population in a given month. The vertical line at "0" indicates January 2003, the exact month in which the new regime started. The graph shows a sharp and substantial decline in DI application rates at the cutoff of about one-third of the pre-GKP level. This effect is only slightly smaller than the effects found by Godard et al. (2024) who only consider permanent workers.

Figure 3: Regression Discontinuity plot for monthly DI application cohort rates, linear fit[a]



[a]: Monthly observations represent application cohorts centered around January 2003.

Table 2: RDiT estimates of the GKP reform on monthly DI application rate

	Dependent Variable DI Application
	Coeff. [% change] (se)
RDiT estimate dropping those within one month of the cutoff	–0.0003*** [33.25%]
24 months bandwidth (=preferred model)	(0.0001)
Mean of dependent variable	0.0009
Number of individual-month observations	260,783,292
RDiT estimate dropping those within one month of the cutoff	–0.0004*** [46.18%]
15 months bandwidth	(0.0001)
Mean of dependent variable	0.0009
Number of individual-month observations	147,399,252
RDiT estimate dropping those within two month of the cutoff	–0.0004*** [36.83%]
24 months bandwidth	(0.0001)
Mean of dependent variable	0.0010
Number of individual-month observations	249,444,888
RDiT estimate dropping those within four month of the cutoff	–0.0003*** [31.62%]
24 months bandwidth	(0.0001)
Mean of dependent variable	0.0009
Number of individual-month observations	226,768,080

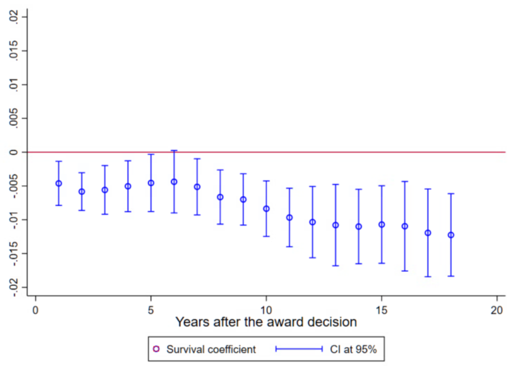
Notes: (1) Standard errors are clustered at the month of the year level; (2) Estimates in brackets are presented in percentage changes of the baseline DI application. (3) ***/**/* = Significant at the 1%/5%/10% level.

Table 2 shows the corresponding RDiT estimates using Equation (1) for our preferred model specification, as well as specifications with different donut sizes and bandwidths. This specification uses a 24-month bandwidth on each side of the threshold with a one-month bandwidth. The preferred specification implies a 33% decrease in DI applications due to the GKP reform. For a smaller bandwidth of 15 months, the effects are slightly larger. Note that our specification controls for calendar-month-fixed effects, which puts constraints on the minimum bandwidth size. Appendix Table C1 complements our findings with RDiT estimates that are stratified by four broad application diagnoses used by the NSII: musculo-skeletal disorders, mental disorders, cardiovascular disorders and "other". The drop in applications is most substantial for mental conditions (-41.3%), followed by musculo-skeletal disorders (-36.0 %), conditions that are typically "hard-to-verify" in assessments (see Godard et al. (2024)).

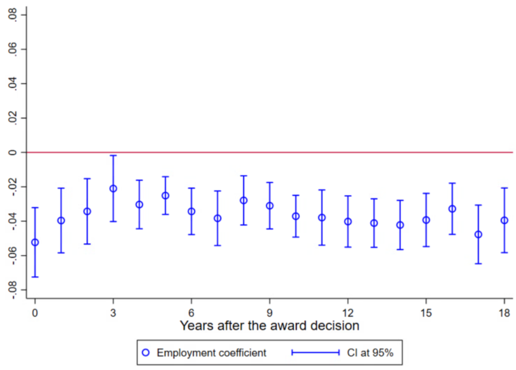
4.3 Effects on survival rates and employment

To see how the GKP reform affects the composition of the pool of applicants, we estimate Equation (1) for survival rates and employment outcomes on the sample of workers who applied for DI benefits between 2001 and 2004. Figure 4 presents the annual estimates up to 18 years after application.[13] To measure elapsed years since application, we use the month of the GKP reform (January 2003) as the reference point for all cohorts. As a result, elapsed times naturally differ across monthly applicant cohorts within a given calendar year. For example, outcomes for the October 2002 applicant cohort are observed about three months earlier than the cohort at the cutoff month. These timing differences are accounted for in the RDiT specification, which flexibly captures any underlying trends around the cutoff date.

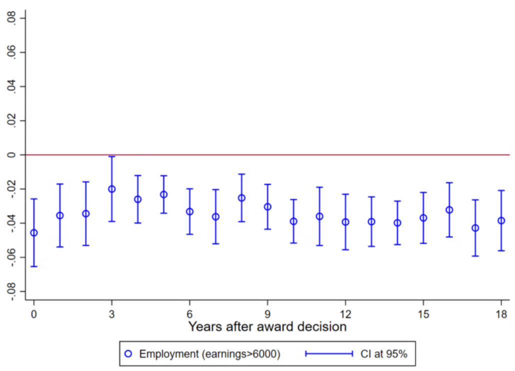
Figure 4: RDiT estimates of GKP reform on survival, employment and DI receipt[a]



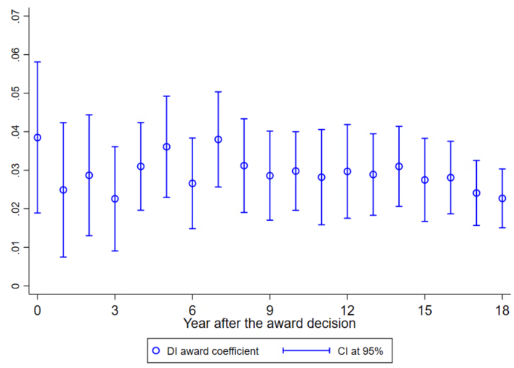
(a) Survival (fractions)



(b) Employment (fractions)



(c) Annual earnings > 6,000 euros (fractions)



(d) DI receipt (fractions)

[a]: The survival rate estimates in panel (a) use the full sample of DI applicant cohorts between 2001 and 2004. The estimates on employment and DI outcomes in panels (b), (c) and (d) are on samples of surviving applicants below the age of 65 up to 18 years after the award decision.

[13] Appendix Tables C2 and C3 provide coefficients for the effects on survival and employment 1, 5, 10, and 15 years after application.

Figure 4a shows statistically significant mortality effects in the years following DI application, amounting to a 1.1 percentage-point difference in survival rates 15 years after the award decision. This corresponds to about 12.2% of the average mortality rate in the pre-GKP cohort. The immediate drop of roughly 0.5 percentage points in the first year after the award decision suggests that workers with short life expectancies were not screened out during the DI waiting period. In addition, Figure 4b shows that the persistent decline in survival rates among DI applicants is accompanied by a reduction in average employment rates. The employment rate falls by about four percentage points in the first year after DI application and remains roughly constant up to 18 years after application.

Applicants in our sample may have experienced some loss of work capacity, but they may still be able to work fewer hours or in jobs with lower hourly wages. We therefore also examine the effects of the GKP reform on earnings among those who remain employed and earn more than 6,000 euros per year. Figure 4c shows a decline in earnings, confirming that the reform reduced hours worked and/or hourly pay. Finally, Figure 4d reports the effects of the GKP reform on the DI receipt rate. We find that post-GKP application cohorts are more likely to receive DI benefits. Since the reform did not alter the DI eligibility criteria but instead only strengthened the screening before application, this result imlies (and consistent with expectations): those who continue to apply after the reform exhibit lower survival and employment outcomes. The estimated effects range from two to four percentage points over the years since DI application.

4.4 Inferences on targeting

Although our results point to persistent and statistically significant negative survival and employment effects of the reform, we need a broader context to characterize these changes as small or large. For improvements in targeting efficiency, compliers to the reform—those discouraged from applying—should have substantially higher future survival and employment rates than the non-compliers who still apply after the reform. To get more insight into this we characterize compliers to the GKP reform, using a simple non-parametric method proposed by Almond and Doyle (2011). In our context, this requires two key assumptions. First, *monotonicity*, meaning that the reform induced workers to stop applying but did not induce any worker to apply who would not have applied before the implementation of the reform. In light of the high DI inflow rates prior to the reform, this assumption is plausible (Koning and Lindeboom, 2015).

As a second assumption, the reform should not have *behavioral effects* on the future outcomes of stayers. In other words, workers who were not deterred from applying (non-compliers) should not experience causal changes in their employment or health outcomes as a result of the GKP reform. Because the DI award rate is higher among post-GKP applicant cohorts than among pre-GKP cohorts (see Table 1), this is a stronger assumption. We address this concern with a robustness test in Appendix D, where we re-estimate our model outcomes on the sub-sample of DI awardees. For this sample, the absence of changes in work incentives is more credible. The results suggest that behavioral responses induced by the GKP reform play at most a limited role.

When both assumptions hold, we can infer the average outcome of compliers to the reform. For this we define $\overline{Y}_{NC}$ and $\overline{Y}_{C}$ as the averages of outcome Y for non-compliers ("stayers") and compliers to the reform, respectively. In addition, we define π as the fraction of the number of compliers relative to the total number of compliers and stayers. β_Y denotes the RDiT coefficient for Y before and after the reform. Following Almond and Doyle (2011), we then can write:

$$\overline{Y}_{C} \quad = \quad \overline{Y}_{NC} - \frac{\beta_Y}{\pi}, \tag{2}$$

To infer the targeting properties of the reform, we compare the implied values of $\overline{Y}_{C}$ of compliers to those of the stayers (non-compliers), $\overline{Y}_{NC}$. Equation (2) shows that the difference between the two averages follows from the RDiT estimate on Y and the fraction of compliers, π.

We can broaden the analysis by considering the outcomes of workers who did not apply in the pre-GKP period: the "non-applicants", $\overline{Y}_{NA}$. Like the applicant cohorts, we also observe outcomes of non-applicants in the years following the reform. Higher targeting efficiency is associated with compliers having outcomes that are close to those of non-applicants, whereas outcomes closer to those of non-compliers indicate weaker targeting efficiency. In line with this logic, we construct the following measure of the degree of targeting for the GKP reform:

$$DT_Y \; = \; 1 \; - \; \frac{\overline{Y}_{NA} - \overline{Y}_{C}}{\overline{Y}_{NA} - \overline{Y}_{NC}} \tag{3}$$

Perfect targeting implies that those in good health (or employment) are screened out (i.e., $\overline{Y}_{NA}$ $= \overline{Y}_{C}$). In that case, the degree of targeting DT_Y equals 1. DT_Y will be zero when the compliers have equal outcomes to non-compliers that kept on applying for DI (i.e., $\overline{Y}_{NC} = \overline{Y}_{C}$) .

Table 3 shows the implied targeting effects DT_Y that follow from equations (2) and (3) for the long-term survival rates (upper panel) and employment outcomes (lower panel). The table also includes (implied) fractions of awarded and rejected stayers (i.e. non-compliers). The upper panel of Table 3 shows that in the first year after the award decision, survival rates of compliers are virtually identical to those of non-applicants. So, in the short run healthy workers are screened out, and targeting is (close to) perfect (DT_Y=1). During this period, the survival rate of compliers to the reform closely matches that of non-applicants. However, in later years survival rates of non-applicants and compliers start to diverge and survival rates of compliers move closer to those of the "stayers", in particular rejected stayers. The result is a substantially lower targeting measure (DT_Y). The conclusion is that it is important to not only focus on short run targeting effects, but that for a more balanced view one should also look at the longer term effects on targeting.

Regarding employment outcomes, the lower panel of Table 3 shows a decline for both stayers and compliers over the years following application, while the employment rate of compliers is closer to that of rejected stayers. Both groups have substantially lower employment rates than the non-applicant group, resulting in low targeting measures DT_Y.

Table 3: Implied targeting effects of the GKP reform

	within one year	**within 5 years**	**within 10 years**	**within 15 years**	**within 18 years**
Survival rate					
Stayers (%)	98.20	95.28	92.06	88.25	85.21
Rejected stayers (%)	99.55	98.22	96.12	93.33	90.87
Awarded stayers (%)	97.45	93.65	89.83	85.44	82.08
Compliers (%)	99.72	96.80	96.91	91.58	88.85
Non-applicants (%)	99.74	99.04	97.99	96.27	94.77
Degree of Targeting	0.98	0.40	0.40	0.41	0.38
Employment rate					
Stayers (%)	59.08	54.71	48.92	50.02	50.38
Rejected stayers (%)	69.06	70.31	61.60	61.61	61.20
Awarded stayers (%)	53.54	45.71	41.13	42.29	42.87
Compliers (%)	66.35	64.41	59.22	62.75	61.59
Non-applicants (%)	95.18	85.41	81.45	80.84	81.62
Degree of Targeting	0.20	0.31	0.32	0.41	0.36

Notes: Survival rate averages apply to all DI applicants between 2001-2004. Employment averages apply to DI applicants still alive and (still) below the age of 65. With the RDiT estimates on DI applicants and on survival and employment, we use Equation (2) to obtain outcomes of compliers.

5 Screening on health or work disability?

Thus far, there are two important findings to note. First, the emerging picture is that the GKP reform induced self-screening of (potential) applicants before or during the DI waiting period. These effects are stronger for workers with mental and musculo-skeletal disorders (see section 4.2), conditions that are typically hard-to-verify. Second, the previous section showed almost perfect targeting on survival in the first year after the GKP reform (i.e., the targeting measure DT_Y being very close to one). But the targeting on survival worsened (DT_Y around 0.4) in later years. This underlines the importance of also evaluating targeting effect of DI reforms in the longer run. As regards employment, targeting seems to be poor in both the short- and the long-run and employment rates of compliers and stayers are substantially lower than the employment rate of non-applicants. This motivates a further look as to how reform-induced changes in health conditions can account for the observed declines in survival and employment outcomes of applicants in later years after the reform. Using information on chronic conditions from the medical prescription database and diagnoses recorded by NSII examiners at the award decision, we therefore construct two indices of applicants' health and work disability: the Survival Health Index (SHI) and the Work Ability Index (WAI). We use these indices to characterize the applicant groups that were screened out by the GKP reform.

5.1 The Survival Health Index (SHI)

Survival rates are often used as a comprehensive and readily observable proxy for health—see Gelber et al. (2023) and Black et al. (2024) for recent contributions. However, they have an important limitation. Substantial differences in health conditions at the time of DI application that persist over time may significantly affect *expected* mortality at later ages, yet for younger workers, the consequences of these conditions for future mortality are often not yet observable and may therefore go undetected. To address this concern, we follow Danesh et al. (2024) by constructing a Survival Health Index (SHI), a composite measure that unpacks the long-term mortality implications of chronic conditions observed at the time of DI application. Similar to Danesh et al. (2024), the projections we use are based on the (pooled) sample of applicants of 65 for whom the consequences of chronic health conditions on survival are more often observed. In our application, we thus predict the 10-year survival rate of *all* applicants based on their current chronic conditions observed in the medical prescription database and application diagnoses of NSII assuming they are still present at age 65.

The construction of the SHI score consists of two steps. First, we select all workers in the DI applicant cohorts who become 65 in some year between 2006 and 2012.[14] For this sample we estimate a linear probability model for the 10-year survival rate S of a worker i (i = 1,..N) in year t aged 65 as a function of (22) chronic diseases CC, and (14) application diagnoses, $DIAG$, and μ as year-fixed effects:

$$S_{i,t} \ = \beta_0 + \beta_I \, DIAG \ _i + f(CC_{i,t})^* \, \beta_{CC} + \mu_t + \epsilon_{i,t} \qquad\qquad \text{if} \qquad AGE_{i,t} = 65 \qquad (4)$$

with ϵ as i.i.d. with an expected value equal to zero and variance $\sigma^2_{\S}$. $f(CC_{i,t})^*$ denotes the set of chronic conditions and all its possible interactions of individual i at time t that are found to contribute to the explanation of the 10-year-survival rate at the age of 65. Since we pool the data, we assume that the effects of the chronic conditions on the 10-year survival rate at age 65

[14]Recall that we only observe medical prescriptions between 2006 and 2022.

remain constant over the years. We estimate Equation (4) using Lasso with CC and $DIAG$ as focal variables.

As a second step, we predict SHI scores for all applicant cohorts aged under 65 from 2001–2004, observed between 2006 and 2022. This setup naturally entails sample attrition due to mortality and workers reaching age 65. The resulting SHI score is given by:

$$SHI_{i,t} \;=\; \hat{\beta}_0 + \hat{\beta}_I \, DIAG_{\,i} + f(CC_{i,t})^* \, \hat{\beta}_{CC} + \hat{\mu}_t \qquad \text{for} \;\; t = 2006, \ldots, 2022 \tag{5}$$

5.2 The Work Ability Index (WAI)

To examine how changes in chronic diseases affect workers' employment outcomes, we complement the SHI with the Work Ability Index (WAI). Like the SHI, the WAI is constructed using the same set of chronic conditions (CC) and application diagnoses ($DIAG$). Our goal is to derive index scores for all workers under age 65,[15] so we predict WAI scores only for applicants up to that age.[16]

Using current employment outcomes as dependent variables in Lasso specifications, the estimation of the WAI scores is more straightforward than for the SHI scores. Specifically, we now use the full sample of workers for which we observe CC in any given year between 2006 and 2022, and estimate models for each year in our sample. Consequently, our model coefficients also vary for each sample year. Specifically, for each calendar year t, we estimate employment E of worker i in year t (t = 2006, .., 2022) who is below the age of 65:

$$E_{i|t} \;=\; \gamma_0 + \gamma_I \, DIAG_{\,i} + g(CC_{i,t})^* \, \gamma_{CC} + \nu_i, \tag{6}$$

with ν that is i.i.d. with expected value zero and variance σ^2_ν. The function $g(CC_{i,t})^*$ denotes the union of chronic diseases and their relevant cross-products.

Following from Equation (6), the WAI is defined as the predicted value of E of individual i in a given year t:

$$WAI_{it} \;=\; \hat{\gamma}_0 + \hat{\gamma}_I \, DIAG_{\,i} + g(CC_{i,t})^* \, \hat{\gamma}_{CC} \tag{7}$$

[15] This reflects the fact that DI eligibility ends at the statutory retirement age of 65.

[16] Sample attrition due to mortality and aging will affect the WAI scores to some extent, but these effects are very similar for the pre- and post-GKP cohorts.

5.3 The impact of conditions and application diagnoses on SHI and WAI

Following the methodology outlined above in Sections 5.1 and 5.2, we construct the two indices—the Survival Health Index (SHI) and the Work Ability Index (WAI)—to assess whether the GKP reform improved the targeting of the DI program. Specifically, we examine the extent to which the reform ensured that DI benefits were allocated to workers with severe health conditions ("health disability") and/or those with low earnings capacity ("work disability").

Figures 5a and 5b present the lasso coefficients from the estimation of Equations (4) and (6), which include the focal variables—chronic conditions and diagnoses of applicants—respectively.[17] For expositional purposes, Figure 5a includes only chronic conditions with prevalence rates above 5%.[18] The residual category "other impairments" serves as the reference group. The SHI coefficients in Figure 5 represent the effects of chronic conditions and application diagnoses on the 10-year survival rate at age 65. For WAI scores, the coefficients capture the effects on the employment rate in the year 2013 (10 years after application).[19]

Figure 5a shows substantial variation in the Survival Health Index (SHI) and the Work Ability Index (WAI) across chronic conditions.[20] The two indices appear to be approximately linearly related; coefficients with high impacts on the SHI score generally also have high impacts on WAI scores and vice versa. Parkinson's disease is the most detrimental condition for both SHI and WAI, reflecting both low survival and low work ability, whereas rheumatic diseases appear to be the least harmful in terms of health and work disability. Hyperlipidemia, acid-related diseases, and psychological disorders do not substantially reduce life expectancy, but have pronounced negative effects on work ability. By contrast, respiratory diseases and diabetes have a large negative effect on life expectancy, but limited impact on work ability.

Turning to the effect of application diagnoses, Figure 5b also reveals substantial heterogeneity in the effects on overall health and work ability. The association between SHI and WAI effects appears weaker than that observed for chronic conditions. Importantly, mental and musculo-

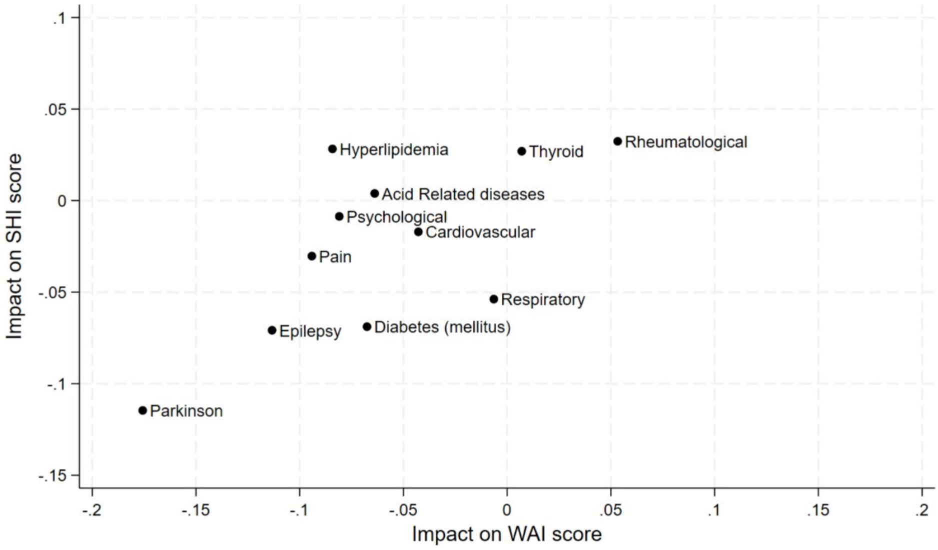
[17] Recall that our models also incorporate cross-products, which are omitted here for expositional clarity. The accumulated effects of comorbidities are not shown.

[18] Appendix Figures A6 and A7 show *all* coefficients for SHI and WAI scores from chronic conditions and diagnoses obtained from Equations (5) and (7) (estimated in 2013).

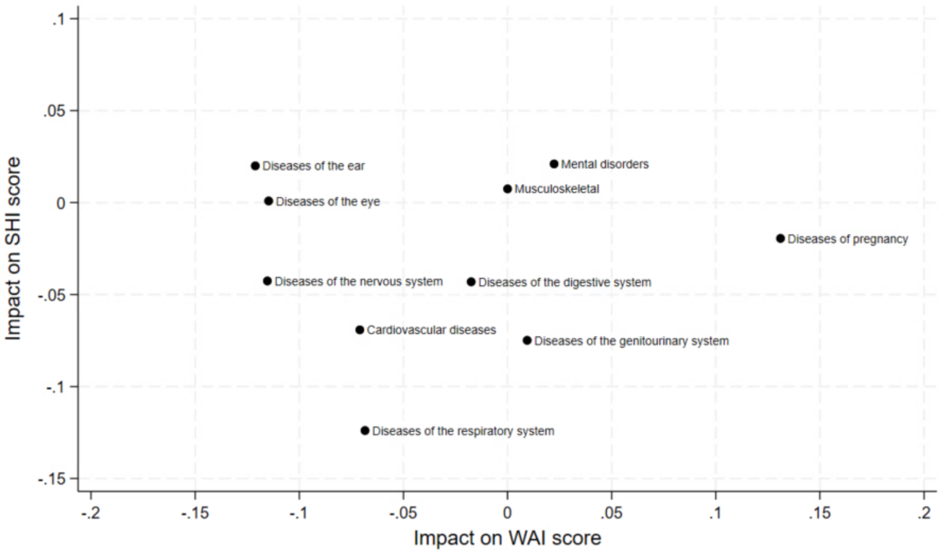
[19] It should be noted that the applicant sample also consists of individuals with conditions that cannot be linked to medical prescriptions in our data. It would therefore be misleading to characterize the baseline group with an impact equal to zero consists of individuals without any health conditions.

[20] To gauge the overall spread and correlation of the indexes, Appendix Figure A8 presents a binscatter of SHI and WAI scores of 2013 of application cohorts. In addition, Appendix Figure A9 displays histograms of both indices.

Figure 5: Lasso coefficients of chronic diseases (a) and medical diagnoses at application (b) as focal variables for the SHI and WAI



(a) SHI and WAI coefficients of **chronic diseases**; year=2013



(b) SHI and WAI coefficients of **application diagnoses**; year=2013

skeletal diseases have no strong effect on survival and employment ten years after application.[21] Diseases of the ear, eye, and nervous system exhibit strong negative effects on the WAI score but not on the SHI score. Respiratory diseases are associated with lower SHI scores.

[21] Note that no LASSO coefficient was obtained for musculo-skeletal diseases as a focal variable in the WAI regression for 2013. Since the explanatory power of the corresponding dummy is limited, we set its value equal to zero in Figure 5b.

5.4 Health disability and work disability

We define applicants with a health disability as having a low SHI score compared to their WAI score, and applicants with a work disability as having a low WAI score compared to their SHI score. Following from this, we construct combined terciles of SHI and WAI scores to distinguish nine different types of disability, as shown in Table 4. We next compress these nine combinations into five main applicant groups: three diagonal groups—(i) high health and high work ability (top tercile for both, labeled as "none" disabilities); (ii) medium health and medium work ability (middle tercile for both, labeled as Middle group); (iii) low health and low work ability (bottom tercile for both SHI and WAI, labeled as both disabilities)—plus two off-diagonal groups: (iv) predominant work disability (upper triangle); and (v) predominant health disability (lower triangle). The pre-reform fractions of these five groups all range around 20% of the total pre-reform sample in 2013.

Table 4: Classification of disability types from SHI and WAI terciles[a]

High SHI (3^{rd} tercile)	Work disability (19.1%)	Work disability	"None" disabilities (22.3%)
Medium SHI (2^{nd} tercile)	Work disability	Middle group (17.6%)	Health disability
Low SHI (1^{st} tercile)	Both disabilities (21.7%)	Health disability	Health disability (19.4%)
	Low WAI (1^{st} tercile)	**Medium WAI (2^{nd} tercile)**	**High WAI 3^{rd} tercile)**

[a]: Fractions of total population of application cohorts (2001-2004) measured in 2013 are shown in parentheses.

Figure 6 presents the five disability groups by their average number of chronic conditions in panel (a) and by the prevalence of the most common application diagnoses in panel (b). As expected, applicants with both disabilities (low SHI and low WAI) have substantially more chronic conditions than those in the Middle group, who in turn have more comorbidities than applicants in the "None" group (high SHI and high WAI). More strikingly, applicants with predominant work disability exhibit about three times as many comorbidities as those with predominant health disability (2.4 versus 0.8). This pattern suggests that predominantly work-disabled applicants are more likely to suffer from multiple chronic conditions that markedly reduce work capacity while having a limited effect on life expectancy. We can also see this reflected in Table 3, discussed in Section 4, showing substantial differences in employment rates of “stayers” versus non-applicants. Appendix Figure A10 provides further details for applicants with a predominantly work disability and a predominantly health-related disability. Among the former, psychological disorders, acid-related diseases, and rheumatological diseases are more prevalent.

Panel (b) of Figure 6 shows that more than 80% of applicants with cardiovascular diagnoses fall into the group having both health and work disabilities, that is, those with SHI and WAI scores in the lowest terciles. Among the two largest application diagnosis groups that are also more difficult-to-verify (mental and musculoskeletal disorders), the picture is very different. Mental disorders are primarily concentrated in the middle SHI/WAI group and among applicants with predominantly work disability. For musculoskeletal disorders, we also see a substantial share of applicants with predominantly work disability, but an even larger share in the group with high SHI and WAI scores (“None disabilities”). The residual category, which includes endocrine disorders, cancers, blood diseases, and other specific conditions, is overrepresented among applicants with predominant health disability. Applicants with these diagnoses therefore tend to face shorter life expectancies but less severe limitations in work capacity.

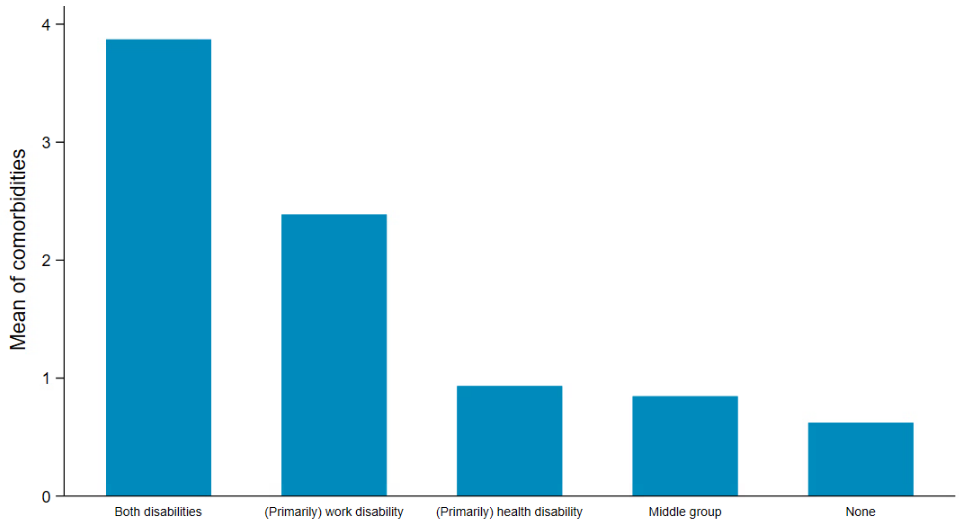
5.5 GKP reforms effects on health and work disability of applicants

With SHI and WAI scores capture substantial variation in survival and employment outcomes, and considering the marked differences between applicants with health versus work disabilities, we now examine which types of workers were screened out most by the GKP reform. Specifically, we assess reform effects on (i) SHI and WAI scores, and (ii) the shares of the five disability categories. Consistent with our earlier analyses, we employ the RDiT model with the same functional form, donut hole, and bandwidth specifications.

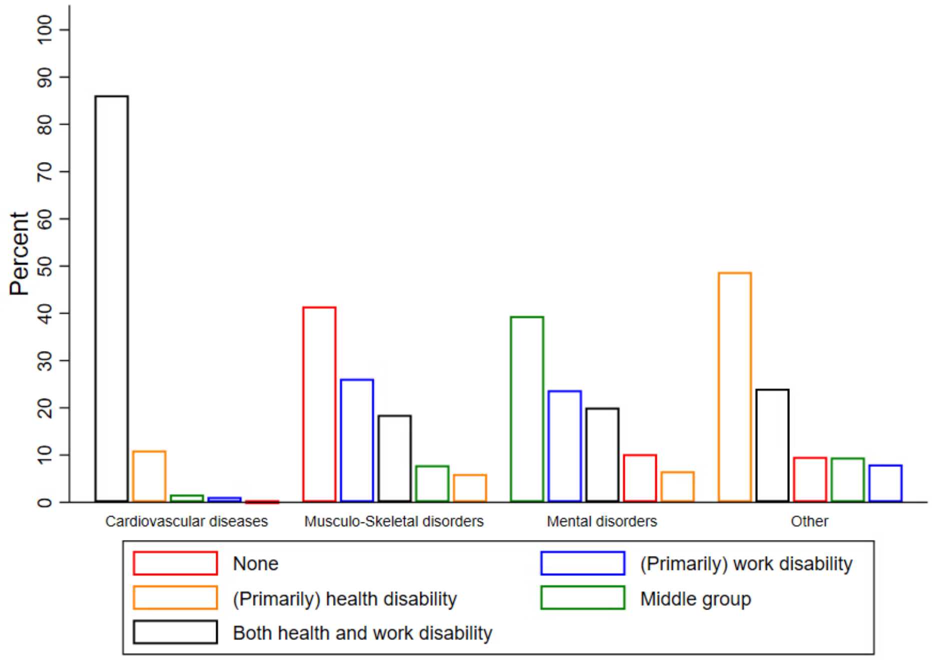
Figure 7 presents the RDiT estimates (in blue) of the reform effects on SHI and WAI scores.[22] For reference, we also include the ‘raw’ RDiT estimates (in red) of the 10-year survival rate for applicants turning 65 between 2006 and 2012 (panel a) and the employment rate of applicants (panel b); these estimates were shown earlier in Figure 4. Since the 10-year survival rate is estimated from a pooled sample, the corresponding -2.4 percentage-point estimate cannot be linked to specific years.

[22]Note that chronic condition data are only available from 2006 onward, explaining the absence of effect estimates for the first three years after application.

Figure 6: Number of comorbidities and application diagnoses by disability category[a]



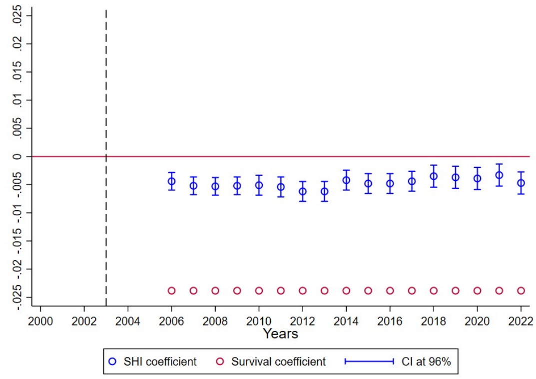
(a) Number of comorbidities



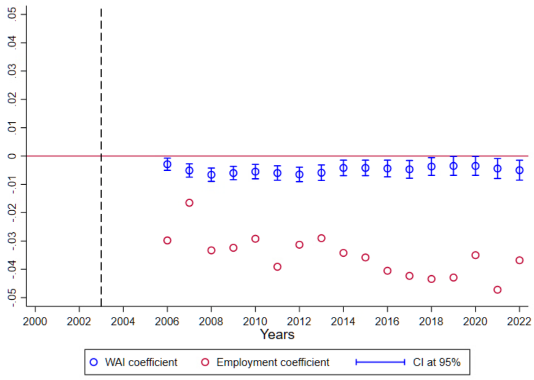
(b) Application diagnoses (fractions)[b]

[a]: For the classification of disability categories of the applicant cohorts of 2001-2004, we use the SHI and WAI scores of 2013; [b] "Other" application diagnoses include cancers, blood diseases and endocrine diseases.

Figure 7: RDiT estimates[a] of GKP reform on Survival Health Index (SHI) and Work Ability Index (WAI) (2006-2022)



(a) Survival Health Index (SHI)[b]



(b) Work Ability Index (WAI)

[a]: Given that chronic conditions are available since 2006, the estimates start three years after the award decision.
[b]: The 10-year survival coefficient in panel (a) shown in red is derived from pooling all observations of individuals at the age of 65 in a given year.

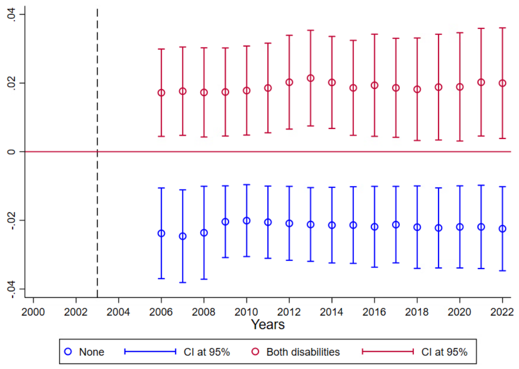
For the SHI, the estimated reform effect in the RDiT design is remarkably stable at approximately -0.5 percentage points over the years. This pattern suggests that differences in later-life survival outcomes between the pre- and post-GKP cohorts emerge as early as the first year after application and persist over time. Compared with the 2.5% effect on 10-year survival reported in red in Figure 4a, this means that about one-fifth of the change can be explained by differences in application diagnoses and chronic conditions.[23] The WAI effects of the reform range from -0.5 to -0.9 percentage points between 2006 and 2022, indicating a decline in employment ten years after the award decision that is attributable to differences in application diagnoses and chronic conditions. Given that the total employment effect, shown in red in Figure 4b, is around 4%, also, only a small share can be explained by the health proxies. This is especially true for the reform effects 15 years after application, where the estimates become borderline significant.

We next examine reform-induced changes in the shares of the five disability categories. Figure 8 presents the RDiT estimates for all five disability categories up to 18 years after the reform. Panel (a) combines the categories “No disabilities” and “Both disabilities” in a single figure, revealing striking mirror-image effects: the share of applicants with health and work disabilities increased by approximately two percentage points (10%), while the share with no disabilities declined by a similar amount. This pattern suggests that healthier workers were screened out, leaving a higher share of applicants with more severe health conditions affecting both survival

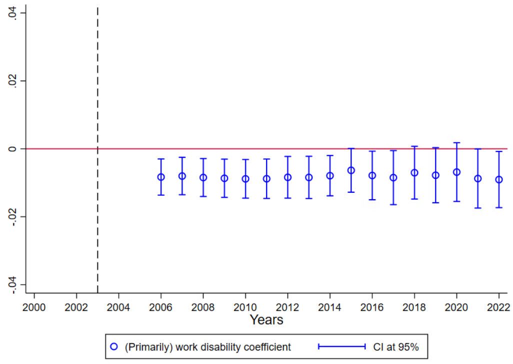
[23]Other important drivers are age and gender.

and work capacity. This implies that targeting efficiency improved in both dimensions.

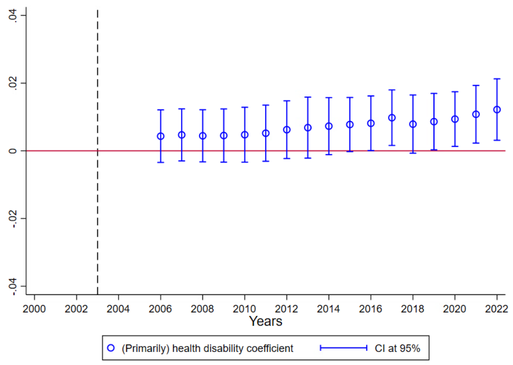
Figure 8: RDiT reform effects on fractions of disability categories[a]



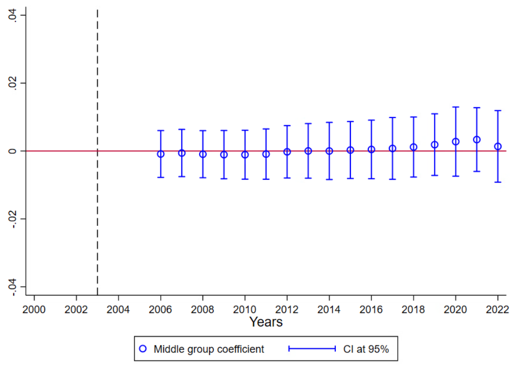
(a) None and both disabilities



(b) Work disability



(c) Health disability



(d) Middle group

[a] We select individuals from the 2001-2004 applicant cohorts who are alive and below the age of 65. Based on these samples, we construct terciles needed for the five disability categories. Note that chronic conditions are observed for 2006 onwards, explaining missing estimates between 2003 and 2005.

Figure 8 also shows an increase in the share of applicants with predominantly health-related disabilities, but the effect does not become statistically significant until after 10 years and reaches about one percentage point after 18 years. This suggests that applicants with predominantly severe health conditions that reduce life expectancy were screened out less than those with milder conditions. The increase in health-disabled applicants is mirrored by a corresponding decline in the share of applicants with predominant work disability, with both categories shifting by roughly one percentage point in opposite directions and with similar precision. Accordingly, the reform also led to a shift in targeting from health-related disabilities to work-related disabilities.

6 Conclusion

This paper examines the long-term health and employment effects of the 2003 reform that introduced stricter screening of disability insurance (DI) applications in the Netherlands. In January 2003, the Gatekeeper Protocol (GKP) substantially increased the obligations of sick-listed workers and their employers during the waiting period prior to DI application. The reform led to a marked decline in DI applications, particularly among workers with hard-to-verify conditions such as mental and musculo-skeletal disorders. These changes raise important questions about the targeting properties of the reform. As the composition of medical conditions among applicant cohorts changed, how were long-term health and employment outcomes affected? What were the outcomes for individuals who, following the reform, no longer applied for DI benefits? And how do compliers and non-compliers ("stayers") compare to non-applicants within the broader working population?

To address these questions, we exploit the sharp cutoff introduced by the GKP reform. Using a Regression-Discontinuity-in-Time (RDiT) design, we estimate that the reform reduced DI applications by approximately 33%. This substantial decline is accompanied by lower survival rates and reduced employment up to 18 years after application among the post-reform applicant pool. Applying the framework of Almond and Doyle (2011), we interpret these patterns as evidence that individuals discouraged from applying by the reform (compliers) exhibit higher survival and employment rates than non-compliers. In particular, the outcomes of the compliers closely resemble those of pre-reform applicants who were denied benefits. Taken together, these findings suggest that the targeting of the DI screening system improved to some extent following the reform, both in terms of health-related and work-related disability. However, these aggregate effects do not reveal which changes in the composition of health conditions among applicants drive the results, nor how such shifts translate into long-term health and employment outcomes.

We delve deeper into medical application diagnoses from assessors from the National Employee Insurance Agency (UWV) and detailed information on chronic diseases to explain changes in long-term survival and employment outcomes. For this we construct a Survival Health Index (SHI) and a Work Ability Index (WAI) to predict as proxies for applicants longer term survival and work outcomes. We find that application diagnoses and chronic diseases may have different impacts on the SHI and WAI scores. Psychological conditions, acid-related diseases and hyperlipidemia are not associated with relatively low residual life time (i.e. lower SHI scores), but substantially reduce the ability to work of applicants (i.e low WAI score). In contrast, respiratory disorders reduce life expectancies but have limited impact on the ability to work.

Finally, we define applicants into five main groups based on the terciles of the SHI and WAI distributions. A striking finding is then that applicants with primarily work disabilities have substantially higher comorbidity rates than those with primarily health disabilities. So, workers with work disabilities typically have multiple chronic conditions that hinder employment, but that do not have strong implications for their survival rates. Consistent with our earlier analyzes, we also find that the reform reduced the share of applicants in relatively good health (i.e., high SHI and high WAI), along with an increase in the share of applicants with both health and work disabilities (low SHI and low WAI). In addition, we find that applicants with predominantly health conditions that substantially reduce life expectancy but have limited impact on work ability were less likely to be screened out than those with milder conditions but low work ability. Thus, the reform disproportionately screened out individuals with predominantly work disabilities, as compared to those with predominantly health disabilities.

Although these results suggest that the overall targeting properties of the Dutch DI screening system improved following the GKP reform, the findings are more nuanced for workers with mental and musculo-skeletal health conditions, the largest applicant groups, for whom the decline in DI applications was most pronounced. Although this group consists of people with health conditions that are often considered difficult to verify by medical assessors, they also often face multiple comorbidities that can substantially impair their ability to work.

These patterns raise broader questions about the design of DI systems. In many countries, increasing DI flows among workers with mental health conditions have prompted calls to tighten eligibility criteria, particularly for conditions that are harder to verify. Although such reforms may be effective in reducing inflow, they also risk screening out individuals with genuinely limited work capacity due to comorbidities, potentially exposing them to significant income losses in the absence of benefit coverage (Deshpande and Lockwood, 2022; OECD, 2022).

Bibliography

Ahammer, A. and A. Packham (2025): "Disability Insurance Screening and Worker Outcomes," *Journal of Health Economics*, 101, 1–14.

Almond, D. and J. J. Doyle (2011): "After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays," *American Economic Journal: Economic Policy*, 3, 1–34.

Benítez-Silva, H., R. Disney, and S. Jiménez-Martín (2014): "Disability, capacity for work and the business cycle: an international perspective," *Economic Policy*, 25, 483–536.

Black, B., E. French, J. McCauley, and J. Song (2024): "The effect of disability insurance receipt on mortality," *Journal of Public Economics*, 229, 1–19.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017): "Rdrobust: Software for Regression-discontinuity Designs," *The Stata Journal: Promoting Communications on Statistics and Stata*, 17, 372–404.

Croda, E., J. Skinner, and L. Yasaitis (2018): "The health of Disability Insurees: An international comparison," Working Paper 2018:28 Department of Economics, University of Venice "Ca' Foscari".

Danesh, K., J. Kolstad, W. Parker, and J. Spinnewijn (2024): "The Chronic Disease Index: Analyzing Health Inequalities Over the Lifecycle," National Bureau of Economic Research (NBER) Working Paper No. 32577.

De Jong, P., M. Gielen, and V. Haanstra-Veldhuis (2014): "Verzekeringsgraad kleine werkgevers," APE report number 1209.

Deshpande, M. and Y. Li (2019): "Who Is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, 11, 213–248.

Deshpande, M. and L. M. Lockwood (2022): "Beyond Health: Nonhealth Risk and the Value of Disability Insurance," *Econometrica*, 90, 1781–1810.

Garcia-Gomez, P. and A. C. Gielen (2017): "Mortality effects of containing moral hazard: Evidence from a disability insurance reform," *Health Economics*, 27, 606–621.

Gelber, A., T. Moore, Z. Pei, and A. Strand (2023): "Disability Insurance Income Saves Lives," *Journal of Political Economy*, 131, 3156–3185.

Godard, M., P. Koning, and M. Lindeboom (2024): "Application and Award Responses to Stricter Screening in Disability Insurance," *Journal of Human Resources*, 59, 1353–1386.

Haller, A., S. Staubli, and J. Zweimuller (2024): "Tightening Disability Screening Or Reducing Disability Benefits? Evidence and Welfare Implications," *Econometrica*, 92, 79–110.

Hausman, C. and D. S. Rapson (2018): "Regression discontinuity in time: Considerations for empirical applications," *Annual Review of Resource Economics*, 10, 533–552.

Huber, C. A., T. D. Szucs, R. Rapold, and O. Reich (2013): "Identifying patients with chronic conditions using pharmacy data in Switzerland: an updated mapping approach to the classification of medications," *BMC Public Health*, 13, 1030.

Johansson, P., L. Laun, and T. Laun (2014): "Screening Stringency in the Disability Insurance Program," *B.E. Journal of Economic Policy and Analysis*, 14, 1–19.

Koning, P. and M. Lindeboom (2015): "The rise and fall of disability insurance enrollment in the Netherlands," *The Journal of Economic Perspectives*, 29, 151–172.

Koning, P. and H. Vethaak (2021): "Decomposing employment trends of disabled workers," *BE Journal of Economic Analysis and Policy*, 21, 1217–1255.

Liebert, H. (2019): "Does external medical review reduce disability insurance inflow?" *Journal of Health Economics*, 64, 108–128.

Maestas, N., K. J. Mullen, and A. Strand (2013): "Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt," *The American Economic Review*, 103, 1797–1829.

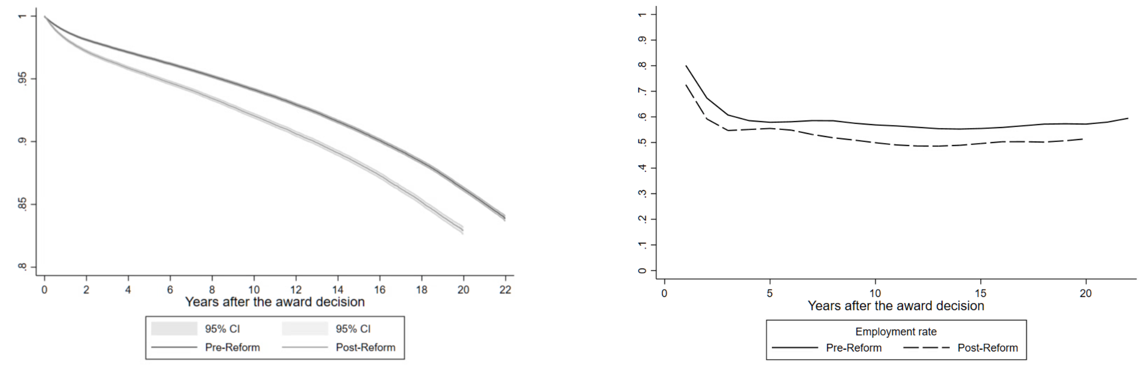
OECD (2010): "Sickness, Disability and Work: Breaking the Barriers. A Synthesis of Findings Across OECD Countries," OECD.

——— (2022): "Disability, Work and Inclusion. Mainstreaming all Policies and Practices," OECD.

Parsons, D. O. (1991): "Self-screening in targeted public transfer programs," *Journal of Political Economy*, 99, 859–876.

A Online Appendix : Figures

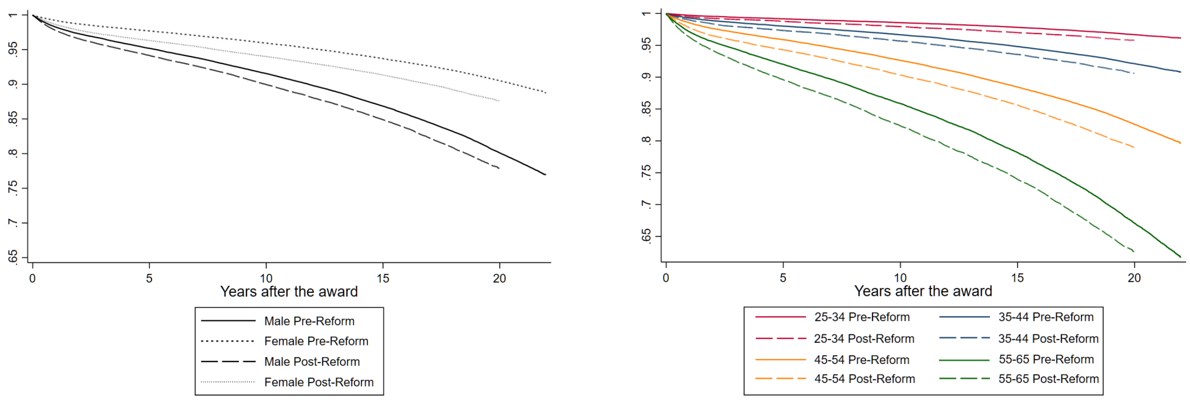
Figure A1: Kaplan-Meier survival estimates and employment outcomes[a] of pre- (2001-2002) and post-GKP reform (2003-2004) DI applicant cohorts



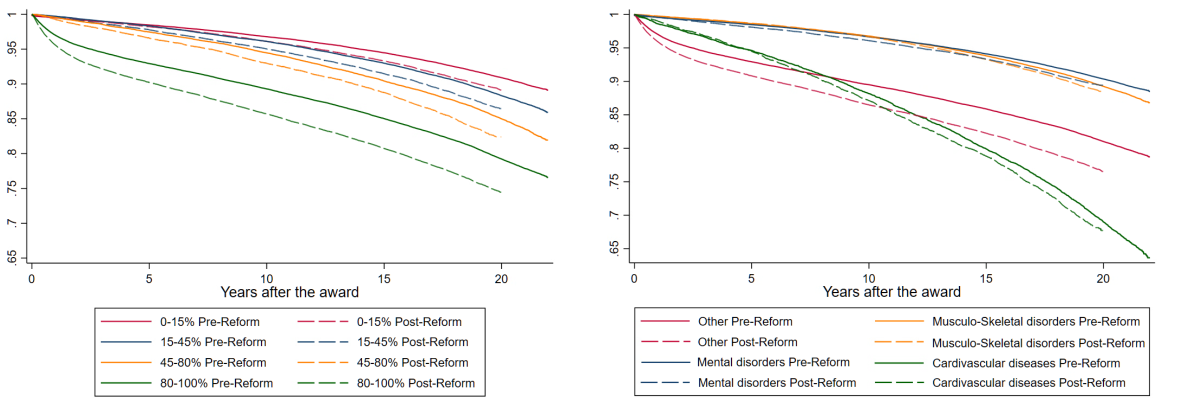
(a) Survival rates of applicant cohorts (b) Employment rates of applicant cohorts

Notes: [a] Employment outcomes are taken over applicants still alive and below the age of 66 in the year after application.

Figure A2: Pre- (2001-2002) and post-reform (2003-2004) Kaplan-Meier survival estimates by DI applicant cohorts stratified by gender, age cohort, degree of disability, and type of impairments



(a) Gender (b) Age categories (at application)



(c) Degree of disability (d) Application diagnoses

Figure A3: Donut RDiT estimates on DI application rates with varying bandwidths

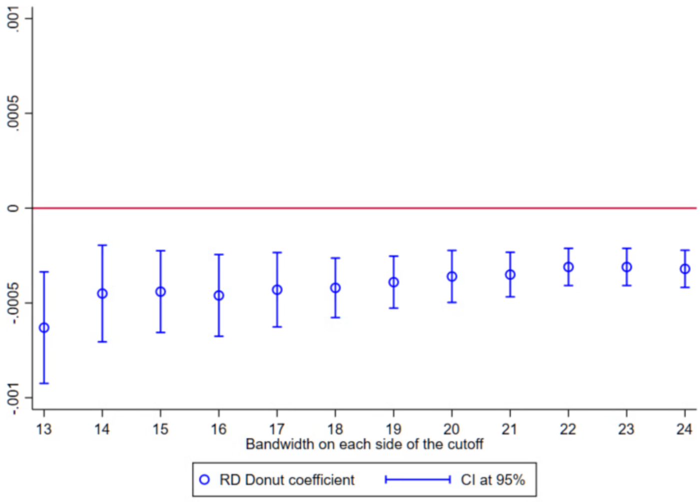
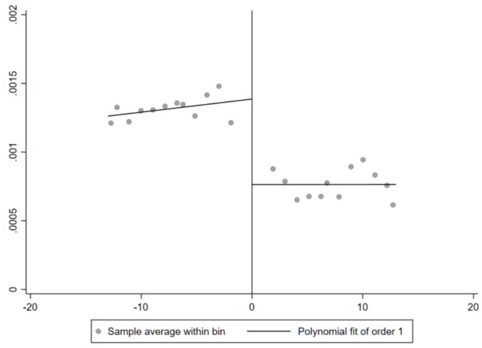
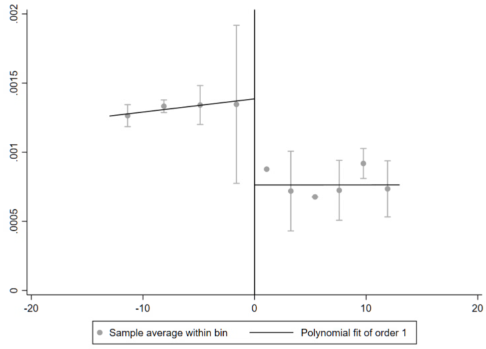


Figure A4: Donut RDiT plots with 13 months bandwidth on each side of the cutoff.



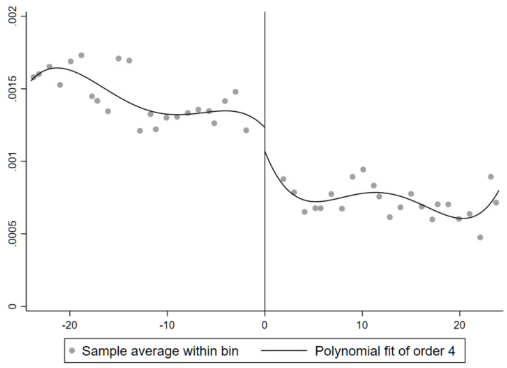
(a) Polynomial of order 1



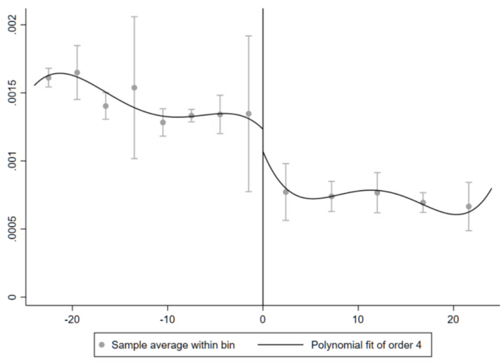
(b) Polynomial of order 1

Notes: (i) Donut RD dropping individuals within one month of the cutoff. (ii) Figure A4b has 95% confidence intervals around the binned means calculated following Calonico et al. (2017).

Figure A5: Donut RDiT plots with 24 months bandwidth on each side of the cutoff.



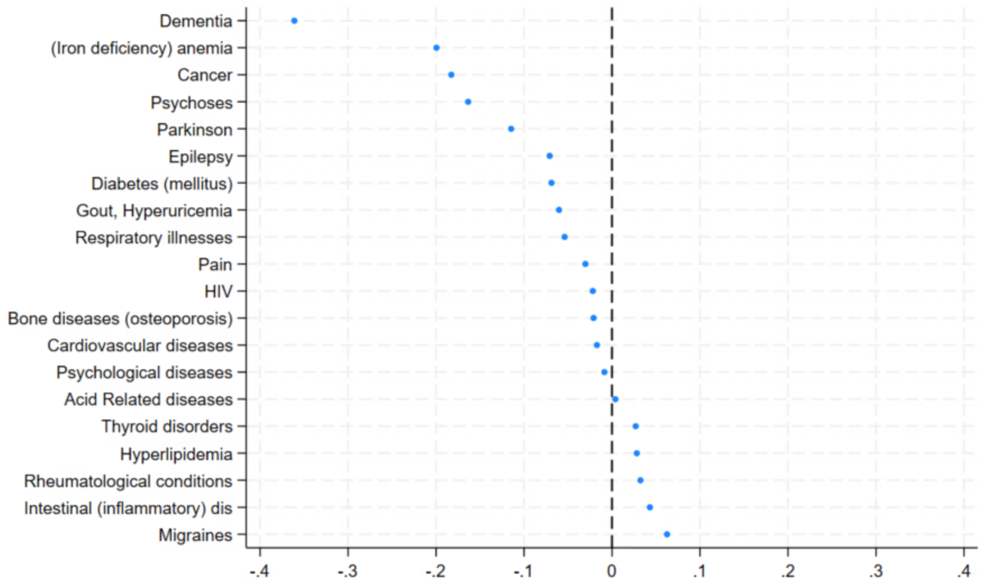
(a) Polynomial of order 4



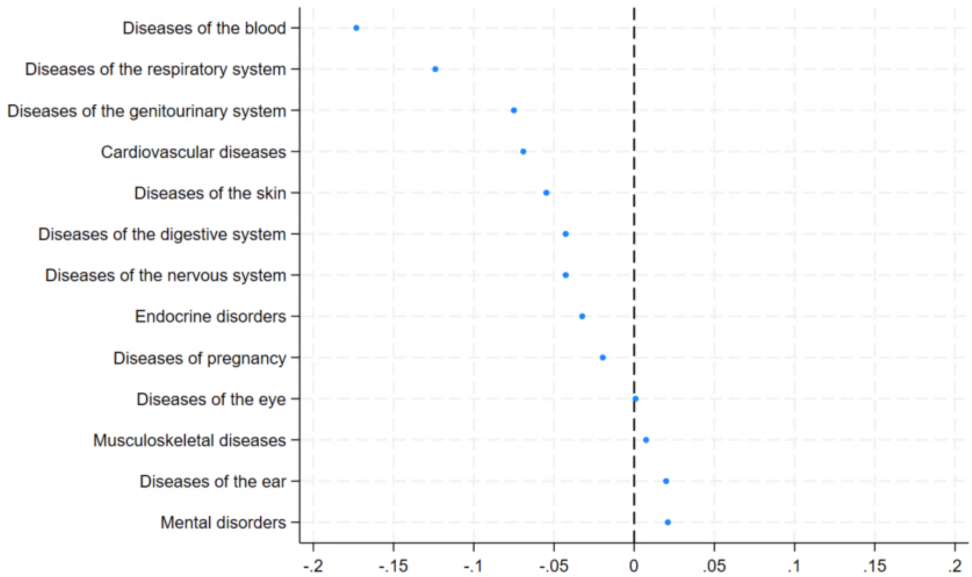
(b) Polynomial of order 4

Notes: (i) Donut RD dropping individuals within one month of the cutoff. (ii) Figure A5b has 95% confidence intervals around the binned means calculated following Calonico et al. (2017).

Figure A6: Lasso coefficients of chronic diseases (a) and and application diagnoses (b) as focal variables for the Survival Health Index (SHI)

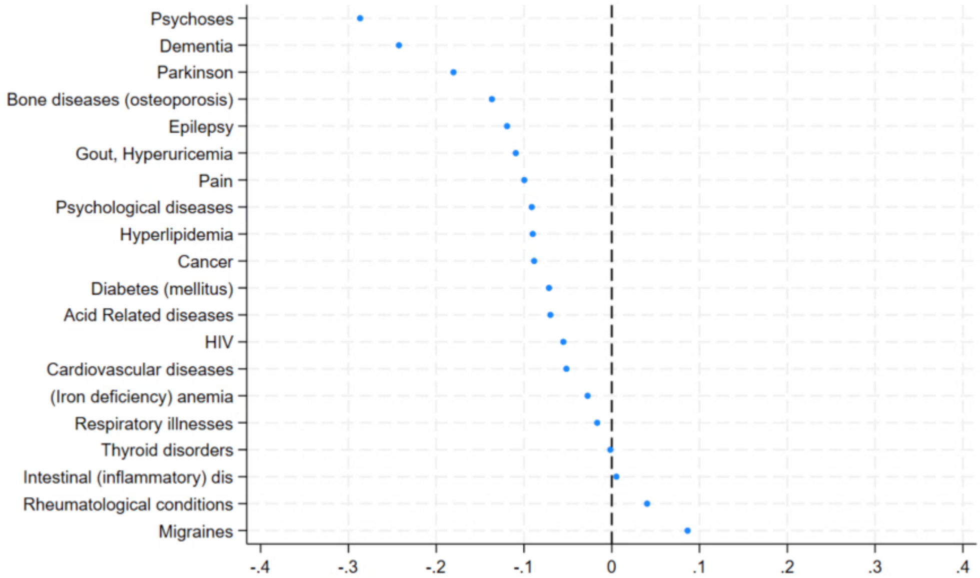


(a) SHI coefficients of chronic diseases; 2006-2012 (65-year old)

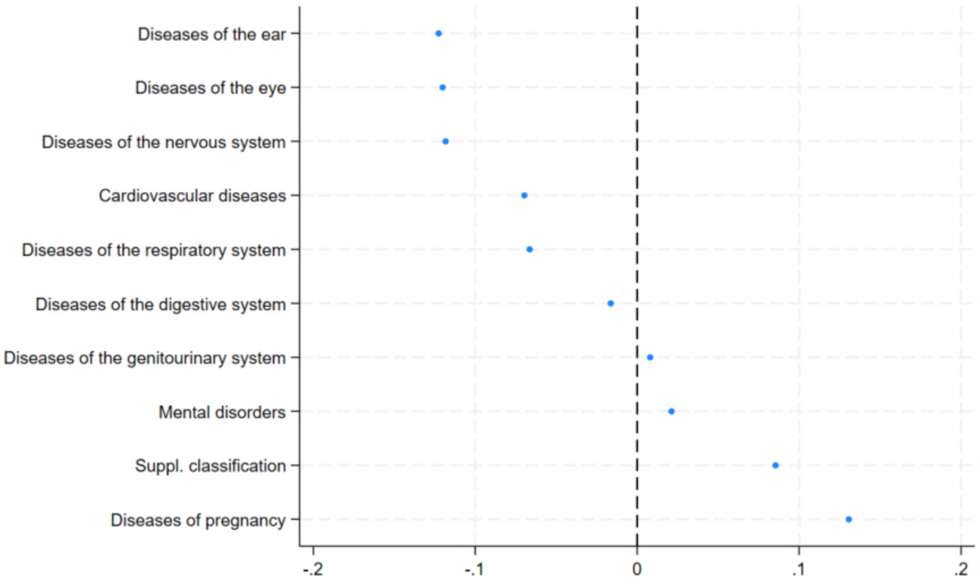


(b) SHI coefficients of application diagnoses; 2006-2012 (65-year old)

Figure A7: Lasso coefficients of chronic diseases (a) and application diagnoses (b) as focal variables for the Work Ability Index (WAI)



(a) WAI coefficients of chronic diseases; year=2013



(b) WAI coefficients of application diagnoses; year=2013

Figure A8: Binscatter of WAI and SHI scores of 2013 (5% percentiles)

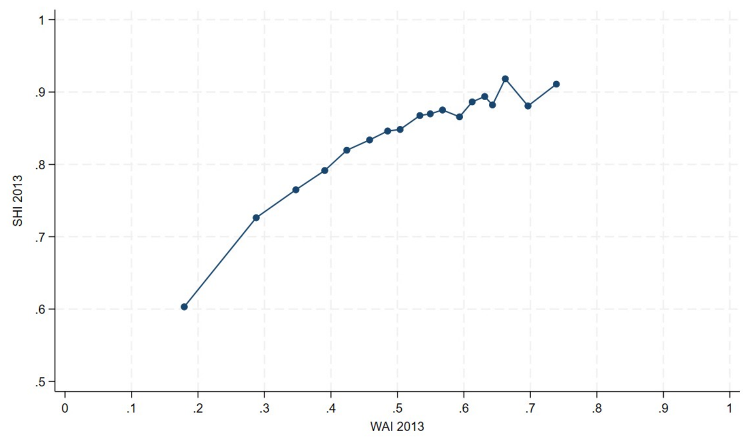


Figure A9: Histogram of SHI and WAI scores in 2013 of applicant cohorts (2001-2004)

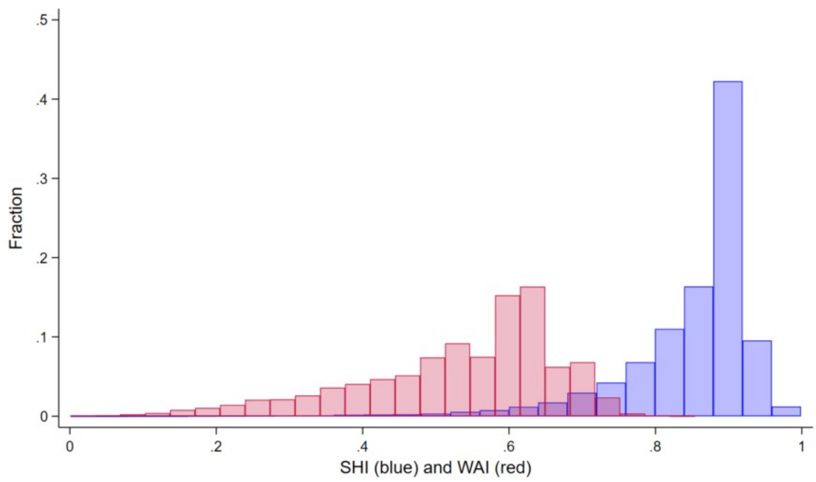
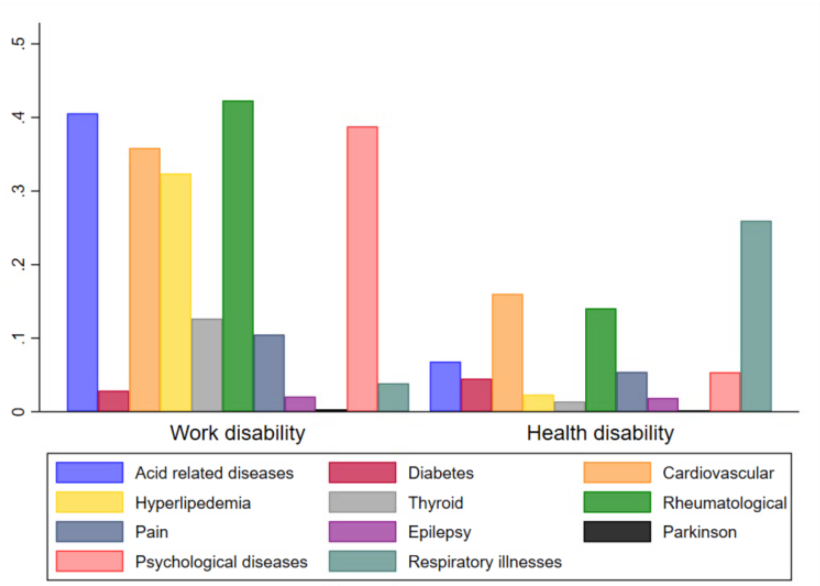


Figure A10: Most prevalent chronic conditions by work and health disability (in 2013)



B Online Appendix : Chronic Diseases

Table B1: Mapping of medical prescription data to chronic diseases

Chronic Disease	**ATC Code(s)**	**Medicine Description**
Acid related disorders	A02	Drugs for acid related disorders
Bone diseases (osteoporosis)	M05	Drugs for treatment of bone diseases
Cancer	L01	Antineoplastic agents
Cardiovascular diseases (inc, hypertension)	B01A, C01, C04A, C02, C07, C08, C09	Antithrombotic agents, cardiac therapy peripheral vasodilators, antihypertensives, beta blocking agents, calcium channel blockers, agents acting on the renin-angiotensin system
Dementia	N06D	Anti-dementia drugs
Diabetes (mellitus)	A10A, A10B, A10X	Insulins and analogues, blood glucose lowering drugs, other drugs used in diabetes
Epilepsy	N03	Antiepileptics
Glaucoma	S01E	Antiglaucoma preparations and miotics
Gout, Hyperuricemia	M04	Antigout preparations
HIV	J05A	Direct acting antivirals
Hyperlipidemia	C10	Lipid modifying agents
Inflammatory diseases	A07E	Intestinal antiinflammatory agents
(Iron deficiency) anemia	B03A	Iron preparations
Migraines	N02C	Antimigraine preparations
Pain	N02A, N02B	Opioids, other analgesics and antipyretics
Parkinson's disease	N04, N05B, N05C	Anxiolytics, hypnotics and sedatives
Psychological disorders	N06A	Antidepressants
Psychoses	N05A	Antipsychotics
Respiratory illnesses	R03	Drugs for obstructive airway diseases
Rheumatological conditions	L04A, M01, M02	Immunosuppressants
Thyroid disorders	H03	Thyroid therapy
Tuberculosis	J04A	Drugs for treatment of tuberculosis

Note: Classification is based on Huber et al. (2013).

C Online Appendix : Tables

Table C1: RDiT Estimates of the GKP reform effect on DI applications by application diagnosis

	Dependent Variable DI Application
	Coeff. [% change] (se)
Panel A: All application diagnoses	–0.0003*** [33.25%]
24 months bandwidth (=preferred model)	(0.0001)
Mean of dependent variable	0.0009
Number of individual-month observations	260,783,292
Panel B: Application diagnoses (preferred model)	
Musculo-skeletal	-0.0001*** [36.00%]
	(0.0001)
Mean of dependent variable	0.0003
Number of individual-month observations	251,237,832
Mental	-0.0001*** [41.30%]
	(0.0001)
Mean of dependent variable	0.0004
Number of individual-month observations	251,860,580
Cardiovascular	-0.0001*** [14.03%]
	(0.0001)
Mean of dependent variable	0.0001
Number of individual-month observations	249,779,126
Other	-0.0001*** [25.50%]
	(0.0001)
Mean of dependent variable	0.0002
Number of individual-month observations	250,123,436

Notes:[a] (1) Standard errors are clustered at the month of the year level. (2) Estimates in brackets are presented in percentage change of the baseline DI application for the sample of interest. (3) ***/**/* Significant at the 1%/5%/10% level.

Table C2: RDiT estimates of the GKP reform on the survival rates of DI applicants

	within 5 years	within 10 years	within 15 years	within 18 years
GKP effect	-0.005*	-0.008***	-0.011***	-0.012**
	(0.002)	(0.002)	(0.003)	(0.003)
Mean of dependent variable	0.966	0.941	0.908	0.883
Number of observations	282,421	282,421	282,421	282,421

Notes:[(a)] (1) Standard errors are clustered at the month of the year level. (2) ***/**/* indicate significance at the 1%/5%/10% level.

Table C3: RDiT estimates of the GKP reform on the employment rate of DI applicants

	within 1 year	within 5 years	within 10 years	within 15 years
GKP effect	-0.039***	-0.025***	-0.037***	-0.039***
	(0.009)	(0.006)	(0.006)	(0.008)
Mean of dependent variable	0.672	0.572	0.503	0.413
Number of observations	279,620	266,459	231,641	184,870

Notes:[(a)] (1) Standard errors are clustered at the month of the year level. (2) *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. (3) Subsample of applicants below 65 years old.

Table C4: Mean prevalence of chronic conditions by application diagnosis (2013)

	Bone dis. (osteo-porosis)	Pain	Migraines	Epilepsy	Parkinson	Psychoses	Psycho-logical dis.	Intestinal inflam. dis.	Dementia	Respiratory ill.	Glaucoma
Cardiovascular	0.025	0.126	0.012	0.060	0.011	0.017	0.131	0.008	0.004	0.154	0.025
Musculo-skeletal	0.023	0.156	0.030	0.044	0.008	0.020	0.147	0.010	0.001	0.145	0.013
Mental	0.016	0.096	0.034	0.043	0.011	0.076	0.277	0.007	0.001	0.129	0.012
Other	0.029	0.116	0.033	0.053	0.013	0.030	0.176	0.014	0.002	0.150	0.018
Total	0.022	0.121	0.031	0.047	0.010	0.042	0.199	0.009	0.001	0.141	0.015

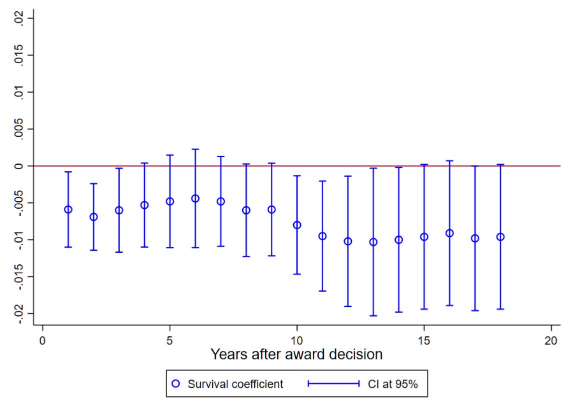
	Diabetes	Iron def. anaemia	Cardio-vascular dis.	Hyper-lipidemia	Thyroid disorders	Tuber-culosis	HIV	Acid related dis.	Cancer	Rheuma-tological	Gout/ Hyper-uricemia
Cardiovascular	0.210	0.024	0.867	0.677	0.045	0.001	0.009	0.367	0.010	0.188	0.059
Musculo-skeletal	0.086	0.016	0.327	0.205	0.044	0.001	0.008	0.256	0.011	0.346	0.020
Mental	0.063	0.018	0.272	0.158	0.049	0.001	0.009	0.202	0.006	0.250	0.011
Other	0.094	0.020	0.339	0.204	0.066	0.001	0.014	0.234	0.011	0.270	0.018
Total	0.086	0.018	0.339	0.211	0.052	0.001	0.010	0.235	0.009	0.281	0.017

D Online Appendix : Estimation results on DI awardees

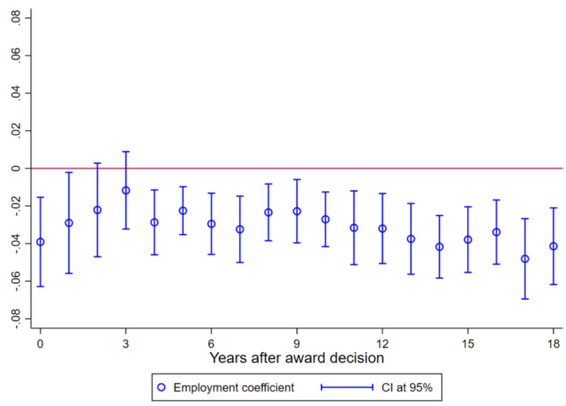
Our inferences on compliers rely on the assumption that no behavioral effects were induced by the GKP reform. Differences in outcomes between pre- and post-GKP application cohorts can then be interpreted as effects of (self-)screening of workers who may potentially apply for DI. This assumption is violated when compositional changes that are induced by the reform go together with higher DI award probabilities. In this respect, Figure 4d indicates that DI award probabilities increased by around two percentage point after the GKP reform. In turn, this may have lowered the (average) incentive to work for the remaining group of applicants.

To address this issue, an obvious way to relax the assumption of absence of behavioral effects is to re-estimate our model on the sub-sample of pre- and post-GKP applicants cohorts who were awarded DI benefits. Figure D1 shows the GKP estimation results that follow from this approach. Although the estimates for DI awardee samples have lower precision compared to the DI applicant cohorts, the effects on survival rates and employment rates are mostly in line with our earlier results for the applicant cohorts. While these results do not provide decisive evidence for the absence of any discouraging effects of higher DI award rates, they suggest that the role is limited.[24]

Figure D1: RDiT estimates of GKP effect on survival and employment on cohorts of DI awardees[a]



(a) Survival (fractions)



(b) Employment (fractions)

[a]: The survival rate estimates in panel (a) are obtained from estimation on outcomes of the full sample of DI applicant cohorts between 2001 and 2004. The estimates on employment and DI outcomes in panels (b) on employment are on samples of surviving applicants below the age of 65 after in the elapsed years since application.

[24]Based on the literature that finds employment reductions of 20 to 30% due to benefit receipt (Maestas et al., 2013), one may argue that the discouraging impact on employment ranges between 0.4% and 0.6%. This confirms the limited importance of behavioral effects on employment.