Screening Disability Insurance Applications and Targeting.*

Mathilde Godard[†] Pierre Koning[‡] Maarten Lindeboom[§]

May 16, 2017

Abstract

JEL code :

Keywords : Disability Insurance, Screening, Targeting Efficiency, The Netherlands.

*We are grateful to Patrick Hullegie for his thorough knowledge of the data, and participants to the CRES-UPF Workshop on Disability Topics in Barcelona and the PSE Labour and Public Economics seminar for helpful comments and suggestions. The authors acknowledge financial support from Health Chair – a joint initiative by PSL, Université Paris-Dauphine, ENSAE and MGEN under the aegis of the Fondation du Risque (FDR). Mathilde Godard acknowledges the support of the EU under a Marie Curie Intra-European Fellowship for Career Development.

[†]GATE-LSE, Univ. of Lyon. Email : godard@gate.cnrs.fr

[‡]Leiden University, VU Amsterdam, IZA and Tinbergen Institute

[§]VU Amsterdam, Tinbergen Institute and HEB Bergen

1 Introduction

The Dutch Disability Insurance (DI) system was internationally known for its extremely high enrolment rates that led some researchers even to classify it as the most out of control disability program of OECD countries (Burkhauser et al., 2008). Indeed, expressed as a percentage of the insured working population, DI enrolment increased rapidly to around 12% in the mid-eighties and then remained more or less constant at this unprecedented level until the beginning of the 21^{th} century (see Figure 1). From then on some radical reforms were implemented that were very effective in curbing DI inflow and DI enrolment. In particular, the figure shows that there has been a spectacular decline in the DI award and enrolment rate since 2002. With DI award rates that went down from 1.5% in 2001 to about 0.5% of the insured population in 2012, DI enrolment has decreased from 11% in 2001 to 7.2% in 2012. Also, spending on disability programs in The Netherlands halved from 4.2 percent of the GDP in 1990 to 2.1 percent in 2007 (de Jong, 2011). This is below the DI spending rates in Sweden (2.2%) and Norway (2.5%). In recent years, the number of DI beneficiaries per worker in the Netherlands has decreased below the level of SSI and SSDI beneficiaries per worker in the US.

It has been argued (see e.g. Burkhauser et al. (2008)) that the introduction of the gatekeeper protocol has been responsible for the huge drop in DI inflow rates in the beginning of this century. The gatekeeper protocol, introduced in April 2002, implied a change in the role of the employer and the National Social Insurance Institute (NSII) in the DI application process. Prior to April 2002, the NSII, the worker and the employer had a joint responsibility to bring a sick listed worker back to work. The law that implemented the gatekeepers protocol stipulated that it was the responsibility of the worker and the employer to undertake measures to increase work resumption rates of sick workers during the sickness waiting period. Next, if work resumption did not occur, the role of the NSII was to act as a gatekeeper and to screen DI applications. The main goal of this reforms was to reduce DI inflow, to increase employment rates of workers with disabilities and to ensure that benefits were provide to those who really needed them. The latter relates to the issue of targeting efficiency. A casual glance at Figure 1 suggests that the reforms were indeed very successful in reducing DI inflow. Less clear is whether the reforms did increase employment rates and improved targeting efficiency.

Increased stringency of the program may on the one hand reduce the number of false positives, but may also increase false rejections and induce part of the truly disabled not to apply. For the latter mechanism we use the term perverse self-screening (Parsons, 1991). This paper looks at these issues. More specifically, we exploit exogenous variation induced by the introduction of the gatekeeper protocol and a nationally implemented field experiment to assess the causal impact of increased screening stringency on DI application behavior. We adapt Parsons' model (1991) for applicant self-screening to the Dutch context and allow a role for different disease types to hypothesize about the dominant self-screening mechanisms among the able and disabled workers when screening stringency is increased and test whether targeting efficiency has improved.

The field experiment used in this paper was also used in previous work (De Jong et al., 2011). Our paper differs in some important ways from this earlier contribution. First, De Jong et al. (2011) only consider short run effects from the field experiment, here we follow applicants (and non-applicants) up to eight years post the year of application. Second, we use linked National Social Insurance Institute (NSII) and Statistics Netherlands data of the entire Dutch population with detailed hospitalization, mortality and labor market information of the full population and diagnosis information of the applicants. Third, we use the theoretical model and the data to identify the effect of increased stringency on self-screening mechanisms among applicant and non-applicant workers groups and the relevance for perverse self-screening effects and the consequences for targeting efficiency.

Our paper also relates to other papers in this area. There is a series of papers that use state level variation in the implementation of DI rules (e.g. Autor and Duggan (2003); De Jong et al. (2011); Gruber (2000); Gruber and Kubik (1997)). Often, denial rates for DI applications are used as a proxy for screening stringency. A potential danger of using observed denial rates as a measure for screening stringency is that, besides the screening policy, they also depend on the composition of the inflow. The composition of the inflow may in turn depend on earlier screening policies. Therefore a high denial rate might reflect a strict screening policy as well as a relatively large fraction of non-eligible workers claiming DI benefits (and hence reflect loose practices in the past). In our analyses we also observe strong changes in the composition of the group of applicants in response to changes in screening stringency. Staubli (2011) uses a change in the Austrian disability insurance program where eligibility criteria for men aged 57 and and older were substantially relaxed. The behavioral response induced by this reform interacts with early retire options via other exit routes like unemployment, social assistance and early retirement of the public pension system [1]. Karlström et al. (2008) use a similar reform in Sweden. To asses the employment effects of SSDI receipt Maestas et al. (2013) compare the medium term (2-4 years) labor outcomes of otherwise similar DI applicants who were allowed or denied benefits only because their application was randomly assigned to different caseworkers with different award propensities. Their analyses effectively relies on within regional/state variation in examiners award propensities. We use exogenous variation induced by a field experiment where we directly controlled the screening stringency of the caseworkers. Caseworkers at local offices of the National Social Insurance Institute (NSII) in two out of 26 regions in the Netherlands were instructed to screen reintegration reports considerably more strictly than elsewhere. The field experiment was implemented at the start of a new DI program in April 2002 where a different screening practice was implemented. The difference in screening practice between treatment and control regions was not announced beforehand. This guarantees absence of any anticipatory responses on the side of the pool of potential DI applicants. We examine the extend to which this variation in stringency influences DI application behavior and infer from longer run (up to 8 years) mortality and labor market outcomes whether targeting efficiency has improves and whether increased stringency may induce perverse self-screening in the sense that disabled decide not to apply.

[1] In Kuhn et al. (2010) it is described how pathways via the unemployment system essentially allowed workers to withdraw from the work force at age 59 and bridge the gap until regular early retirement benefits.

The remainder of this paper is organized as follows : Section 2 provides more information about the DI reform and the field experiment. A stylized model of disability insurance application behavior is presented in section 3. Data are presented in section 4. Section 5 includes the empirical specification of our model and Section 6 provides the estimation results. We end with a discussion and conclusion section (Section 7).

2 Institutional background and the experiment

Institutional context: the Gatekeeper protocol of 2002

This section describes the 2002 reforms and draws heavily on De Jong et al. (2011). Before we proceed by explaining the screening process that precedes DI claims, it is important to note to important features of the Dutch DI system that are different from most other OECD countries. First, the DI program covers all income losses that result from both occupational and non-occupational injuries, providing insurance of 70% of the corresponding loss of income. As Koning and Lindeboom (2015) argue, this broad setup of the scheme has increased the possibility of sizable screening errors in disability determinations. Second, wage payments for sick workers are continued in the waiting period that precedes disability claims and the employer bears the wage and sick-pay costs for this period; this period is also referred to as the 'waiting period'. As a result, the system does not provide strong financial incentives for sick-listed workers to resume work quickly. Taken together, these two characteristics of the Dutch DI scheme laid the ground for a broad coverage of impairments and limited self-screening, thus causing high disability levels.
The Gatekeeper protocol stipulates the responsibilities of the worker and the employer for sickness spells lasting six weeks or longer. After six weeks of absence, a first assessment of medical cause, functional limitations and a work resumption prognosis has to be drafted by a medical doctor from an occupational health service agency. On the basis of these data the employer and employee together draft an accommodation and rehabilitation ("reintegration") plan in which they specify an aim (resumption of work in the current job or somewhere else, whether accommodated working conditions are required etc) and the steps needed to reach that aim. This reintegration plan should be ready by the eighth week of sickness. It is binding for both parties, and one party may summon the other when considered negligent. After 13 weeks of absence the employer should report the sick employee to the NSII. From this moment on the worker is added to the administrative database of the NSII. Figure 2 gives a schematic representation of the different steps taken in this protocol.

The worker can file a DI application if s/he has not resumed work before the expiration of the waiting period. At the start of the Gatekeepers protocol in 2002 the waiting period was one year, in 2004 this was extended to two years. Benefit claims are only considered admissible by the social benefit administration if they are accompanied by the rehabilitation plan and an assessment as to why the plan has not (yet) resulted in work resumption. If this procedure is not followed, the employer is obliged to continue providing sick pay for some additional months rather than having the worker transfer to disability benefits.

This implies a strong financial incentive for employers to focus their attention at the onset of sickness, when the opportunities for recovery and work resumption are probably most substantial. On the other hand, employer and the sick worker are jointly responsible for reintegrating the sick worker back into his or her old job or into a new job commensurate with the worker's limitations. However, if the employee consistently rejects reasonable offers and accommodations, the employer may stop paying sickness benefit and, eventually, may fire the employee.

The field experiment

Since all workers becoming sick after April 1, 2002 were treated according to the gatekeeper protocol and DI applications are made in the 39th week of sickness absence, the gatekeeper protocol started to affect DI claims as of January. This was also the start of our experiment. The experiment ended in October 2003. The NSII has 26 regional offices where the caseworkers screened the incoming DI applications. For 24 out of the 26 regional offices the caseworkers were instructed to conduct the standard procedure. That is, to screen the reintegration reports 'on paper' and to only deviate from this and contact the employer and/or the worker only if there was some suspicion of fraud and/or negligence. In the remaining two regions case workers were instructed to deviate from the standard procedure and to always contact the worker and/or employer unless it was absolutely clear that given the condition sufficient measures and reintegration activities had been undertaken. This holds for instance for the severe, the obviously permanently and fully disabled workers and those who have resumed work for more than 50%.

The experiment was set-up in cooperation with the NSII and it was made sure that the treatment regions were provided additional resources in order to implement the stricter screening policy. Since the regions were assigned to us it becomes important to check pre-experiment differences between the treatment and control regions. We will employ a Differences-in-Differences framework and therefore it is essential that pre-treatment outcomes would at least have moved in a common trend. We turn to that later. It should be stressed once more that the experiment started with the introduction of the new gatekeeper protocol and that therefore none of the case workers at the regional offices of the NSII had any experience in screening reintegration reports. For the validity of the research design it is therefore important to check upon the screening practices in the 26 regions and to see if indeed the two treatment regions conducted a stricter screening procedure than in the control regions. To monitor the screening practices we have sent out questionnaires to all 26 regions. In treatment regions time spent on screening reintegration reports by caseworkers was 40% higher than in control regions. Table 1 reports results on the actual use of various screening mechanisms used in the regions. The table shows that in the treatment regions caseworkers visited sick workers (Apeldoorn region) or their employers (Hengelo region) more often than elsewhere. In the control regions reintegration reports were checked more often only on paper or by phone. Since face-to-face confrontations can be considered as the strictest form of screening, these data confirm that screening was indeed stricter in the treatment regions.

3 Theoretical model

Below we closely follow Parsons (1991), but where necessary adapted to fit the Dutch context. The policy objective is to provide income support to those who cannot engage in substantial gainful activity. Consider a population composed of two types, j=0,1: type 0, the able who can engage in substantial gainful activity and type 1 the disabled who can not engage in substantial gainful activity. Both types incur costs from work. The able have to sacrifice leisure and there may be fixed costs of work such as commuting time, for the disabled work may be taxing and substantial health cost arise. Let CW_j, be the costs for worker type j, with $CW_1 > CW_0$. Of course there might be heterogeneity within each of the worker type groups. There is variation in the relative preferences for income and leisure, fixed costs of work may vary and the disabled may vary in severity and type of diagnosis. Severity and type of diagnosis determine the work capacity. In addition, for some diseases it might be very easy to verify the eligibility status for disability benefits, while for others the classification problem is more complicated as there might be more discretion in the decision to award disability benefits. As an example of the former one could think of certain type of cancers where prospects of survival are extremely low. An example of the latter may be mental conditions or musculoskeletal conditions.

Starting a DI application is costly. It involves direct costs associated with reintegration efforts that have to take place in the waiting period and other administrative costs associated with the application.It can also include foregone promotion options or salary increases during waiting period T. [2] The indirect costs are likely to be higher for able workers. Direct costs of severe and easy verifiable diseases are expected to be lower than for less easy to verify diseases. We denote total cost by C. Benefits B are paid to those who apply for benefits *and* are judged to be eligible. Limited (welfare) benefits may also be paid to those not certified as disabled. Denote this means-tested social minimum as B_m, where $0 \leq B_m < B$.

Health is known to the individual, in contrast to the NSII administrators who have to infer the individual's disability status from ancillary traits and/or tests. For easy verifiable diseases the true disability status can be inferred with little error. Classification errors are bigger for diseases where traits/tests are imperfect indicators of the true disability status. In these cases there is scope for opportunistic behavior. Also, increases in screening stringency are expected to be more effective in such cases. Denote H_j^* as the true health/disability status of a j type worker, with $0 < H_j^* \leq 1$ and where 0 denotes death and 1 perfect health and $H_1^* < H_0^*$. Only $H_{ij} = H_j^* + \eta_{ji}$ is observable by the program administrators, where η is a random variable, with $E[\eta_{ji}] = 0$ and $V[\eta_{ji}] = \sigma_j^2$. Denote the probability of benefit allowance by ϕ_j, where $0 < \phi_j < 1$, j=0, 1.

$$\phi_j = \mathbb{P}(H_{ji} < HT|apply) = F_j(H_j^* - HT|apply) \tag{1}$$

Where HT is the health threshold used by the DI administration in the award decision, and F_j is the

[2] Note that in the Dutch context during the waiting period workers are covered by sickness benefits that fully replace wage earnings.

cumulative distribution of the random term η_j. The variance of η_j is small for easy verifiable diseases and larg(er) for other less easy to verify diseases. Average allowance rates will be higher for the disabled than for the non-disabled $\phi_0 \leq \phi_1$. An active screening mechanism is perfect if $\phi_0 = 0$ and $\phi_{1=1}$.

A worker decides to apply if the expected utility of applying exceeds the expected utility of not applying. Let W_j be potential wage of the jth worker type, then with cost C of entering the application process we can write for the expected utility to apply:

$$E(U_j|apply) = \phi_j(U(B) - C) + (1 - \phi_j)(U(W_j) - CW_j - C) \tag{2}$$

for the expected utility not to apply,

$$EU(U_j|noapply) = U(W_j) - CW_j \tag{3}$$

so that decision to apply follows from,

$$I_j = \phi_j(U(B) - U(W_j) + CW_j) - C \tag{4}$$

It is obvious that application rates will rise when benefits are increased. Of interest for the current paper is that $\frac{\delta I_j}{\delta \phi_j} > 0$, $\frac{\delta I_j}{\delta C} < 0$. For able workers (type 0) increasing screening stringency leads to lower award rates ϕ_0 and higher application costs C (for instance because stricter requirement on the reintegration report requires more activities) and subsequently to lower application rates [3]. Also note that moral hazard will be higher among applications for diseases that are more difficult to verify. We expect therefore that screening stringency will result in stronger reduction in application rates for such diseases.

Working is no option for those with prohibitively high health at work costs CW_1 and therefore their decision to apply is based on a comparison of expected utilities from applying and possibly getting a benefit B:

$$E(U_0|apply) = \phi_0(U(B) - C) + (1 - \phi_0)(U(B_m) - C) \tag{5}$$

Or not to apply and get social benefits B_m:

$$EU(U_0|noapply) = U(B_m) \tag{6}$$

The apply decision then follows from

$$I_1 = \phi_0(U(B) - U(B_m) - C \tag{7}$$

So, the disabled (type 1) may decide not to apply (here referred to as perverse self-screening of the

[3]Note that reintegration activities in the waiting period prior to the application can also result in work resumption. In this case one would also observe a decline in application rates. Therefore, in our context self-screening includes the actual decision not to enter the path of disability insurance application as well as work resumption due to successful reintegration efforts during the waiting period

disabled) if C is very high and/or if ϕ_0 is very low. This may be of particular relevance for diseases that are difficult to verify with observable traits and/or tests. After all, reintegration activities will be much more extensive and award rates much lower than for other disabled with easily verifiable diseases. As a consequence it is expected that also among the disabled increased screening stringency will lead to a stronger decline in diseases that are more difficult to verify.

What to expect from increases in screening stringency?

Increased screening stringency will lead to a reduction in application rates of both the able and the disabled workers. It is also expected that the decline in application rates is largest in disease types that are difficult to verify, such as mental and musculoskeletal disorders (i.e. Easy verifiable diseases are less responsive). Now suppose that targeting efficiency is met and that as a response to an increase in the screening stringency none of the able apply anymore, but all disabled still apply. In such a situation one should observe mortality rates to go up in the pool of applicants. Also, in the pool of non-applicants Unemployment Insurance (UI) and Social Assistance (SA) rates should remain stable. After all, able workers that would previously have been in the pool of applicants now move to the pool of non-applicants. Alternatively, stringency of screening may have gone too far, in the sense that able but also some disabled who need a benefit stopped applying (perverse self-screening of the disabled). In such a situation, mortality rates should increase in the pool of applicants as well as in the pool of non-applicants. Furthermore, among the non-applicants one would expect to see a reduction in employment rates and an increase in UI and/or SA rates.

4 Data

4.1 Presentation of the sample

The data used in this paper come from several administrative data sources. One of the central registers for this study is the *DI application register* maintained by the NSII. This record contains information on all DI applications from 1999 through 2013. For each applicant, it provides information on the type of impairment, as well as the date and result of the decision reached by the NSII. This allows us to identify applicants who were awarded benefits, those whose claims were rejected, and those who recovered before the final examination. If the applicant is awarded DI benefits, we have information on his or her disability category.

We merge the *DI application register* with linked records maintained by Statistics Netherlands. These records cover the entire population during the years 1995-2013. To measure the employment and earnings of individuals, we merge our sample with the tax records register. That information provides longitudinal gross earnings (in 2010 Euros) from 1999 to 2013. Based on that information, we consider a worker employed if he or she has any positive earnings in a given calendar year.[4] The tax register also gives

[4]Note that according to this definition, an employed individual can combine both earnings and DI benefits.

information on whether the individual receives disability, welfare or unemployment benefits in each year. Demographics (month-year of birth, sex, nationality and place of residence) are obtained from the municipality register. For each individual, we derive information on the NSII region in which he or she is currently living. The hospital register (LMR) is used to assess individual health of the total population. The LMR register contains data on both inpatient and day care patients of all general and university hospitals and most of the specialized hospitals in the Netherlands from 1998 to 2005. Note, however, that most hospitalization spells due to mental disorders are not registered in the LMR, as they are typically taken care of in specialized hospitals or psychiatric clinics. For the entire Dutch population we observe (i) whether the individual entered the hospital, (ii) the admission and discharge date, and (iii) the main diagnosis. Admissions are categorized by diagnoses on basis of the International Classification of Diseases (ICD) 9-CM. We use the following categorization of hospitalizations : mental, musculo-skeletal, endocrine, cardio-vascular, nervous, respiratory and other diseases. We generate a set of dummy variables taking value one if the individual was hospitalized (all-cause or cause-specific) in the three years before starting his or her (potential) waiting period (i.e. hospitalized in t-1, t-2 or t-3). We also construct an health index that aggregates information over multiple hospitalization spells since 1995 (see Data Appendix C for more information on the construction of the Charlson Comorbidity Index). The death register is used to identify mortality status (all-cause and cause-specific), which will be our key indicator of targeting efficiency (see Parsons (1991)).

The sample of interest consists of a repeated cross section of individuals over 2001-2004, the pre-gatekeeper protocol period (2001 and 2002) and the first two years of the gatekeeper protocol (2003 and 2004). We restrict our sample to prime-age individuals – aged 25-65 – and only consider individuals who were employed in the previous year. We exclude rejected DI applicants (according to the NSII data) who nonetheless appear as disability recipients in the tax register (11.3% of rejected applicants). Conversely, individuals who are awarded DI benefits according to the NSII data, but who do not appear in the DI records of Statistics Netherlands are excluded (6.3% of awarded applicants). We exclude those who apply more than once during the period under study. Finally, for each year in 2001-2004, we take all individuals who apply that year as well as a 10% random sample of non-applicants that year. Our sample consists of a maximum of 701,191 individuals each year.

4.2 Summary statistics

Figure 3 shows the DI application and award inflow over 1999-2004. There is a striking decline in DI application rates and DI award rates from January 2003, when the first DI applications under the new Gatekeeper protocol arrived at the NSII. The share of prime-age individuals who applied to DI decreased from 1.2% of the population in 2001 to 0.5% in 2004. Similarly, the share of the insured population that started collecting disability payments in the a given year declined from 0.7% percent in 1999 to 0.3% in 2004. Interestingly, the figure also shows that while both application and award rates decline after 2003, award rates decline to much a lesser extent. Whether the decline came from able workers, disabled workers or from both is the question at stake in this paper. However, note that the system became stricter with the introduction of the gatekeeper protocol. The relative small decline in award rates suggests that the average health condition of the applicants has worsened after 2003 and hence that the reduction in DI applications was bigger among able workers than among disabled workers. Seen from another angle, among individuals who stopped applying under the new Gatekeeper regime, some would have applied and would have been awarded disability benefits in the old system.

We start with a first look at this issue of screening efficiency in the raw data. To do this, we first look at the direct impact of the introduction of the new system by examining differences in labour-market, health and subsequent mortality experiences of applicants and non-applicants in the old system (2001-2002) and the gatekeeper system (2003-2004). Table 2 reports the means for the group of applicants by period in columns (1) and (3), for the group of non-applicants we report the means in columns (2) and (4)). Relative to the sample of applicants, non-applicants are more likely to be male, young, employed and have higher earnings and are less likely to be foreign-born. Non-applicants are also significantly in better health (based on past hospitalization status and mortality risk) than applicants. It is worth noting that relative to non-applicants, labor-market and health outcomes of applicants seem to have worsened after the introduction of the Gatekeeper protocol (see column (5)). It suggests that the composition of the applicant pool worsens as screening intensifies. Note also that the award rate (conditional on applying) increases from 60 in 2001-2002 to 62% in 2003-2004, which also suggests that the pool of applicants under the Gatekeeper regime consists of more severe cases. This is in line with our observations regarding Figure 3.

In Table 3, we further investigate the issue of screening efficiency. Are those who do not apply or those who are denied benefits systematically healthier? Does screening efficiency increases upon the introduction of the Gatekeeper protocol? The consequences of increased stringency of screening should also be reflected in subsequent mortality experiences of the applicant and non-applicant groups. Following Parsons (1991) and Treitel (1979), we look at mortality outcomes five years after the initial application. Before the introduction of the Gatekeeper system (see Panel A), the 5-year mortality rate of those who apply is more than three times of those those who do not apply : 3.05 percent versus 0.99 percent. The 5-year mortality rate of those who are awarded DI benefits is almost three times larger than that those who are denied : 4.18 percent versus 1.45 percent. After the Gatekeeper was introduced (see Panel B), mortality differentials increase substantially: The 5-year mortality rate of those who apply is now more than four times larger than that of those who do not apply : 4.14 percent versus 0.99 percent. The 5-year mortality rate of those who are awarded DI benefits is almost four times larger than that those who are denied : 5.67 percent versus 1.50 percent. The mortality rates of the non-applicants and the denied have not or hardly changed after the introduction of the gatekeeper system. Hence, the basis of these simple differences in means, screening efficiency seems to have improved after the introduction of the Gatekeeper protocol.[5]

[5]Tables A1-A4 in the Appendix replicate this exercise for earnings, employment, UI and welfare receipt.

The time variation provided by the introduction of the Gatekeeper system is useful for descriptive purposes, but is not sufficient to establish the causal impact of increases in the screening stringency. For this we in addition use the regional variation induced by the field experiment. Table 4 presents summary statistics of the data using a 10% random sample of the population of applicants and non-applicants. Columns (1) and (2) presents means of treated and controls for the pre-gatekeeper period (2001-2002), columns (3) and (4) presents the means for the gatekeeper period (2003-2004). As mentioned in Section 2 treatment regions were not randomly selected but assigned by the NSII. It is therefore important to check if the choice of treatment regions is unrelated to relevant outcome measures (De Jong et al., 2011). As the experiment started in 2003, we can use the pre-treatment years 2001 and 2002 to investigate random assignment. A simple look at columns (1) and (2) suggests that treatment regions do not differ substantially with respect to the outcome variables from control regions. DI application rates, employment and health characteristics are virtually identical across treated and control regions and t-tests confirm this formally. We conclude that the treatment regions were not selected on their DI application rates (or any other relevant outcome) in 2002. When turning to columns (3) and (4), we see that DI application rate halves between 2001-2002 and 2003-2004. However, the mean difference in difference (column (5)) shows that the decline is not differential with respect to the control and treatment regions. Note also that this does not hold for some of the controls (native) and hospitalization outcomes(musculo-skeletal problems and mental health).

Finally, Table 5 present summary statistics for the sample of applicants (based on NSII data) in treated and control regions, in the pre-gatekeeper and the gatekeeper period. Column (5) presents the difference-in-differences mean. In 2001-2002, around 34% of DI applicants were hospitalized in the past three years. This proportion increases to around 40% in 2003-2004, with the strongest increases is in treated regions (see column (5), a 1.5 percentage points increase in the treated relative to control regions). In 2001-2002, mental disorders accounted for the majority of impairments among DI applicants (37 and 40% for the control and treated regions, respectively). The share of mentally-impaired individuals decreased dramatically after the introduction of the gatekeeper system, again with the strongest decline in the treated regions (-4.5 percentage points relative to the control regions). This is in line with the hypotheses that we formulated in Section 3, where we argue that the response to an increase in screening stringency is biggest for diseases that are less easy to verify. The decreasing share of mental diseases is compensated by an increase in the share of "other" (+1/7%) and musculo-skeletal (+1.6%) impairments within the pool of applicants. Those suffering from mental disorders most often have co-morbidities. The results might therefore suggest that with increased stringency a fraction of those who formerly applied for DI benefits on the gorund of mental impairments now choose to apply for other conditions. The lower panel of the table presents DI award rates (conditional on applying) in treated and control regions, in the pre-gatekeeper and gatekeeper period. The table shows an increase in award rates for mental disorders of 8,4 percentage points and a decerase of 4,7 percentage points for musculo-skeletal problems. This is consistent with our previous finding, that part of those applying shift from mental disorders to other disorders (musculo-skeletal) and that only the very cases remain to apply for mental disorders.

5 Empirical strategy

5.1 The impact of stricter screening on DI application behaviour

Our parameter of interest is the effect of screening on DI applications. We measure the impact of stricter screening during the first two years after it was introduced. As a first step, we estimate the following difference-in-differences model :

$$Y_{rt} = \alpha + \gamma Treatment_{rt} + \delta_t + \lambda_r + \epsilon_{rt} \tag{8}$$

where $Treatment_{r,t}$ is the treatment variable. It is a dummy variable taking value 1 if region r= Apeldoorn/Hengelo *and* if year t=2003/2004, and value 0 else. δ_t and λ_r denote year and region fixed effects respectively. Standard errors are clustered at the regional level.

These estimates do not exploit individual characteristics which are available in the data. Taking account of individual heterogeneity is attractive for two reasons : (i) to reduce the variance in the data and (ii) to investigate whether there is a differential impact of screening stringency for different types of workers (for instance by diagnosis). We estimate the following individual level differences-in-differences model for the probability to applying for DI with OLS :

$$Y_{irt} = \alpha + \gamma Treatment_{r,t} + \delta_t + \lambda_r + \beta X_{irt} + \epsilon_{irt} \tag{9}$$

Our individual regressors X_i include gender, age (in 8 categories) and whether or not the individual was born in the Netherlands (nativity). Standard errors are clustered at the regional level.

5.1.1 Validity of the identification strategy

The key assumption in the difference-in-differences framework is that the outcome in treatment and control group would follow the same time trend in the absence of the treatment. This is the so-called "parallel trends" assumption for which we have to see whether it also holds in our data. A first glance at figure 3 shows that the data do seem appear to move in parallel prior to the introduction of the gatekeeper system. In order to test this more formally, extend our regression model by interacting the treatment variable with yearly time dummies :

$$Y_{irt} = \alpha + \delta_t + \lambda_r \gamma_{-1} Treatment_{rt} + \gamma_1 Treatment_{rt} + \gamma_2 Treatment_{rt} + \beta X_{irt} + \epsilon_{irt} \tag{10}$$

We leave out the one interaction term for the last pre-treatment period (i.e., 2002). All other interactions are expressed relative to the omitted period which serves as a baseline. If the outcome trends between treatment and control groups are the same, then γ_{-1} should be insignificant. An attractive feature of this test is that also the interactions of the time dummies after the treatment with the treatment indicator is informative. For instance, γ_1 and γ_2 show whether the treatment effect fades out over time, stays constant, or even increases.

5.2 Stricter screening & targeting efficiency

Stricter screening may lead to a fall in DI applications (see above). But whether it improves the targeting efficiency of the DI program remains to be seen. As our data contains health and employment information on workers in different regions in different years, our empirical strategy allows us to look at this issue in more detail. As a first step, we test whether the decline in application rates is largest in "able" types (as opposed to "disabled" types). Of course disability type is not observable directly in the data. We proxy it by the individual's hospitalization history (e.g. Charlson index). We then estimate Equation 10 separately for the sub-sample of individuals in poor health (according to our health index) and for the sub-sample of individuals in good health. As a second step, we estimate Equation 10 for the sub-sample of DI applicants. We look at the impact of stricter screening on their employment and mortality outcomes up to seven years after the initial application.[6] We next repeat the same exercise on the sub-sample of non-applicants, and look at their subsequent employment and mortality outcomes. If stricter screening had perverse effects in the sense that truly disabled workers would not apply for DI, we should expect to see a decrease in employment rates an increase in UI and Social Assistance among the non-applicants.

6 Results

Estimates of Equation 8 are reported in Table 6, Table 8 presents the results of equation 9. Year specific treatment effects (equation 10) are given in Table 7. This last table also more formally tests for the common trend assumption. For now we can state that the assumption of common trends between treatment and control regions in the period before the introduction of the gatekeeper system can not be rejected. The treatment effects in Tables 6 and Table 8 are significant at the 10% level and imply a reduction in the application rate of about 4.4%. Note that the treatment effect is unaffected after inclusion of additional controls. This is reassuring as it indicates that the non-random assignment of the treatment regions by the NSII is not affecting our parameter of interest.

Table 7 reveals that the treatment effect is essentially driven by the effect in the first year of the gatekeeper protocol. Note that in this first year non of the involved, workers, employers and gatekeepers had any experience with the new system. Nor was it announced before hand that some regions would impose stricter screening practices. Therefore, most likely, the larger part of the reduction in the number of applications comes from increased work resumption rates due to reintegration activities that the worker and employers had to undertake in the waiting period. Alternatively, some workers may have decided not to pursue further in the application procedure once they found out that what was expected from them (i.e. once they found out that C was higher than expected.

[6] The hospital register does not provide information after 2005, therefore we cannot look at long-run effects in hospitalization outcomes for the group of non-applicants. Mortality is thus our key indicator for targeting efficiency.

We furthermore estimated the model on different subsamples to see whether there are heterogeneous effects with respect to past health status. The first three rows present results of a categorization based on the Charlson Comorbidity Index (CCI), the other rows are indicators based on hospitalization rates in the past three years. The Charlson index is based on a longer history of hospitalizations and therefore may be measured with less error. Lower values of the Charlson index are associated with better health. Table 9 presents the treatment effects of these regressions. Those with a zero health index (the healthy), reduce their applications significantly and (again), the effect is only found for 2003. We find a similar coefficient for the healthy based on (the absence of) hospitalizations. However, the hospitalization rates (based on hospitalization rates form the past 3 years) also show substantial and significant effects for some conditions. For instance, for those with musculo-skeletal and cardiovascular diseases we see substantial reductions in the application rates.

The descriptives in the data section (see Table 5) indicated that applications for menetal conditions fell in response to strciter screening. We repeat the analyses here with regressions diagnosis specific applications. The results are presented in Table 10 and show that the reduction in applications comes primarily from a reduction in applications for mental disorders, accompanied by increases in applications for DI due to muscolo-skeletal disorders and "other" disorders. This confirms our findings of section 4 and hints at substitution between different types of disorders when screening stringency is increased. This could be interpreted as pressure to report other impairments than mental disorders. With increased stringency, caseworkers may be less inclined to award decisions for diagnosis where it is more difficult to verify the severity of the condition. Interestingly, the effect persists in 2004. As argued earlier, the year 2003 was the first year that the new year was implemented and therefore one would expect that this effect may for an important part be driven by successful reintegration measures in the waiting period, rather than self-selection/screening on part of the (potential) applicant. The second year may include such self-screening effects.

The results of Tabel **??** should also lead to a change in the composition of the pool of applicants. The results in Table 11 confirm this. In the gatekeeper period the sample of applicants is generally sicker: in this period, the applicants are more often hospitalized. And in line with the findings above, the share of mental disorders in the pool of applicants decreases substantially, accompanied by almost equal, but opposite change in the share of other diseases. Award rates for the fully disabled have not increased, despite the decline in the average health conditions of the applicants. The confirms that stringency has increased.

With a change in the composition in the pool of applicants the essential question whether this self-screening effect has led to improved targeting efficiency, i.e. whether the decline of the average health condition came from a reduction of the able applicants and whether some of the disabled may have decided also not to apply. We refer to the latter as perverse self-screening of the disabled. Figures 5 and figure 6 of the appendix depict the short and longer-run (up to 7 years) effect of stricter screening on labor and health outcomes for DI applicants and non-applicants, respectively. The dots in the figure refer to Differences-in-Differences estimates of the models for each year after the (year of) applications [7]. The effects thus measure the impact of stricter screening in the treatment regions and complements the results on the five year mortality rate of Table 3 of section 4. There is evidence that mortality rates of the applicants significantly increases in the first year after the application, confirming that the average health conditions has worsened among applicants. For the non-applicants we see a reduction in earnings and employment rates, but these are quantitatively small. Mortality rate increases are small (much smaller than for the group of applicants) and generally insignificant. It might be that the size of the pool of non-applicants is too large and that there is too much variation in the outcome variables for other reasons so that it is very difficult to detect negative effects of stricter screening.

7 Conclusion

The Dutch Disability Insurance (DI) system was internationally known for its extremely high enrolment rates that led some researchers even to classify it as the most out of control disability program of OECD countries (Burkhauser et al., 2008). Recently some radical reforms were implemented that were very effective in curbing DI inflow and DI enrolment. It has been argued (see e.g. Burkhauser et al. (2008)) that the introduction of the gatekeeper protocol has been responsible for the huge drop in DI inflow rates in the beginning of this century. The gatekeeper protocol redefined the role of the National Social Insurance Institute. It stipulated that it should that it was the responsibility of the worker and the employer to undertake measures to increase work resumption rates of sick workers during the sickness waiting period. Next, if work resumption did not occur, the role of the NSII was to act as a gatekeeper and to screen DI applications. The main goal of this reforms was to reduce DI inflow, to increase employment rates of workers with disabilities and to ensure that benefits were provide to those who really needed them. The latter relates to the issue of targeting efficiency. This paper exploits exogenous variation induced by the introduction of the gatekeeper protocol and a nationally implemented field experiment to assess the causal impact of increased screening stringency on DI application behavior. We use linked National Social Insurance Institute (NSII) and Statistics Netherlands data of the entire Dutch population with detailed hospitalization, mortality and labor market information of the full population and diagnosis information of the applicants. We estimate sijmple differences in differences models to identify the effect of increased stringency on self-screening mechanisms among applicant and non-applicant workers groups and the relevance for perverse self-screening effects and the consequences for targeting efficiency.

We find that increased stringency in screening has induced behavioral responses in the group of applicants. There was a substantial reduction in the number of applications after the introduction of the gatekeeper. The field experiment used in this paper also showed that screening stringency had a causal impact on applications. The larger part of this effect came from a reduction in the first year of the program. This led us to conclude that successful reintegration efforts in the waiting period prior to the

[7] We only take treatment in 2003 and compare DI applicants in 2003 versus 2001-2002 in treated versus control regions and look at leads in their mortality and labor market outcomes

actual application in week 39 were most likely responsible for the decline in applications. We also find that in line with our theoretical predictions the decline was biggest in applications for mental impairments, impairments that are not easy to verify without error. The crucial question is whether stricter screening has increased targeting efficiency. We examined mortality rates and labor market outcomes up to 7 years after the year of application. There are elevated mortality risks in the pool of applicants, but importantly, not in the pool of non-applicants. This suggests absence of perverse self screening of disabled workers and improved targeting efficiency.

Bibliography

Autor, D. H. and M. G. Duggan (2003): "The rise in the disability rolls and the decline in unemployment," *The Quarterly Journal of Economics*, 118, 157–206.

Burkhauser, R. V., M. C. Daly, and P. R. De Jong (2008): "Curing the dutch disease: Lessons for united states disability policy," Michigan Retirement Research Center Working Paper 2088-188, University of Michigan.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie (1987): "A new method of classifying prognostic comorbidity in longitudinal studies: development and validation," *Journal of chronic diseases*, 40, 373–383.

De Jong, P., M. Lindeboom, and B. Van der Klaauw (2011): "Screening disability insurance applications," *Journal of the European Economic Association*, 9, 106–129.

de Jong, P. R. (2011): "Sickness, disability and work: Breaking the barriers – A synthesis of findings across OECD countries – By OECD," *International Social Security Review*, 64, 103–104.

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

Gruber, J. and J. D. Kubik (1997): "Disability insurance rejection rates and the labor supply of older workers," *Journal of Public Economics*, 64, 1–23.

Karlström, A., M. Palme, and I. Svensson (2008): "The employment effect of stricter rules for eligibility for DI: Evidence from a natural experiment in Sweden," *Journal of Public Economics*, 92, 2071–2082.

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.

Kuhn, A., J.-P. Wuellrich, and J. Zweimüller (2010): "Fatal attraction? Access to early retirement and mortality," Institute for the Study of Labor (IZA) Discussion Paper no. 5160.

Larsson, L. (2006): "Sick of being unemployed? Interactions between unemployment and sickness insurance," *The Scandinavian Journal of Economics*, 108, 97–113.

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.

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

Stagg, V. et al. (2015): "CHARLSON: Stata module to calculate Charlson index of comorbidity," *Statistical Software Components*.

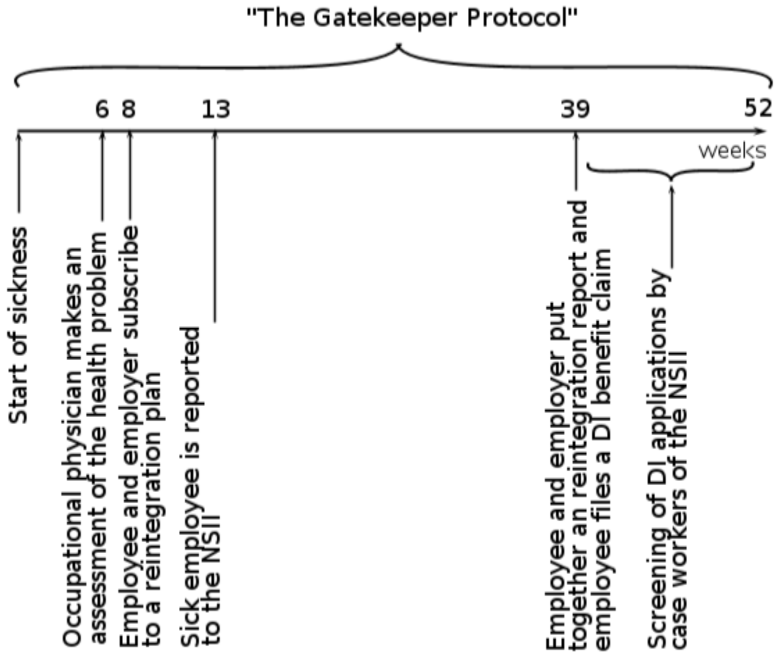
Staubli, S. (2011): "The impact of stricter criteria for disability insurance on labor force participation," *Journal of Public Economics*, 95, 1223–1235.

Treitel, R. (1979): "Disability claimants who contest denials and win reversals through hearings," Tech. rep., Staff Report No. 34, Washington : Soc. Security Admin., Office Res. and Statis.

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

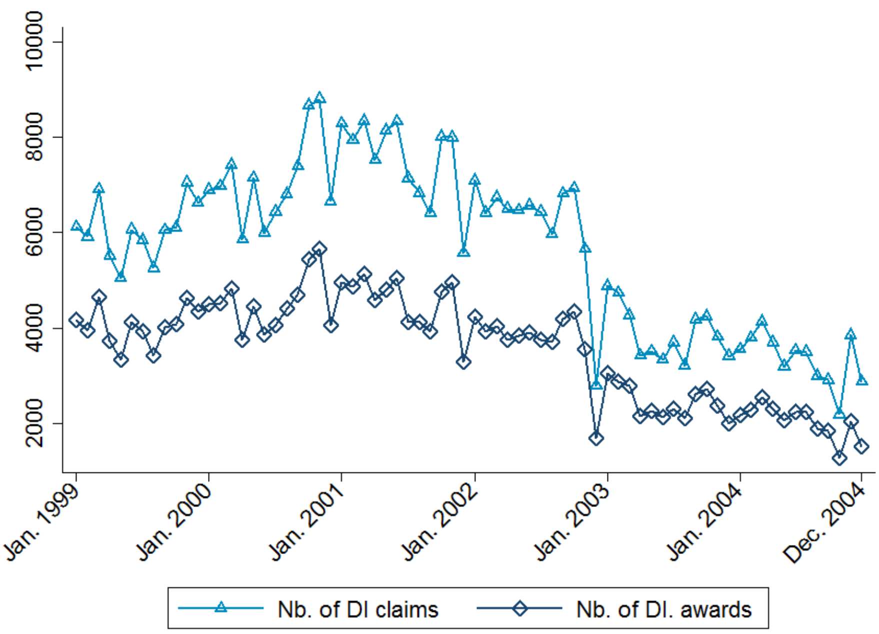
Note : The Disability Insurance award rate is the share of the insured population that started to receive disability payments in a given year. Source : UWV (2012). Borrowed from Koning and Lindeboom (2015).

Figure 2: Schematic representation of the process toward entering DI



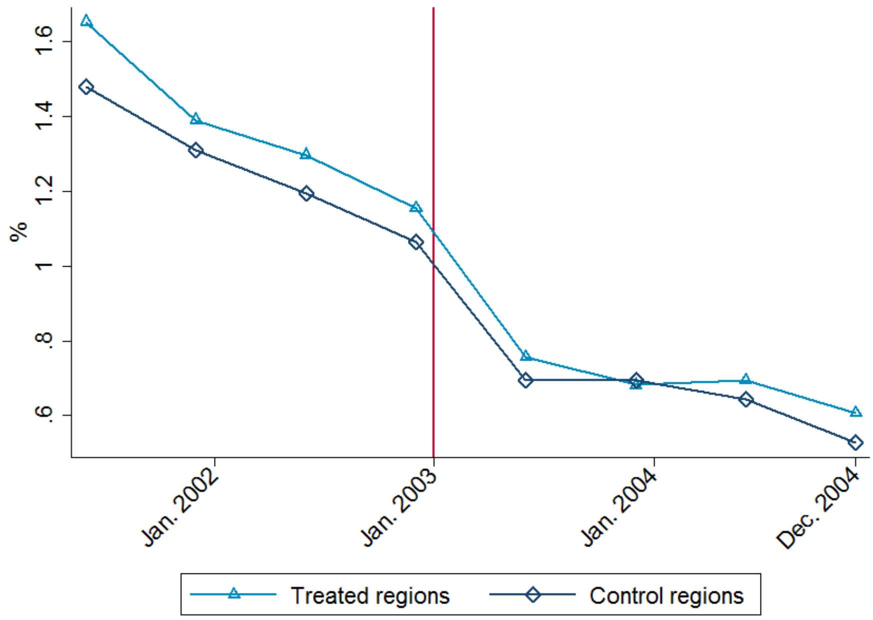
Note : Borrowed from De Jong et al. (2011).

Figure 3: Disability Insurance Application and Award Inflow in the Netherlands, 1999-2004



Note : The DI Application Inflow is the number of prime-age individuals who filled a DI benefit claim in a given year; The DI Award Inflow is the number of prime-age individuals that started collecting disability payments in a given year.

Figure 4: DI application rate in treated and control regions, before and after the experiment.



Note : The vertical red line represents the start of the experiment (January 1, 2003), when the first DI applications under the new Gatekeeper protocol arrived at the regional offices of the NSII.

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%	

Notes : Borrowed from De Jong et al. (2011). Note that caseworkers can use multiple screening methods on one application, so columns can add up to more than 100%.

Table 2: Summary statistics of DI applicants and non-applicants, before and after the introduction of the Gatekeeper protocol.[a]

	2001-2002		2003-2004		Diff-in-
	Applicants	Non-applicants	Applicants	Non-applicants	diff
	(1)	(2)	(3)	(4)	(5)
Demographics					
Age	43.27	41.06	43.77	41.72	-0.17***
	(9.83)	(10.05)	(9.90)	(10.20)	(0.05)
Male	0.41	0.58	0.48	0.56	0.08***
	(0.49)	(0.49)	(0.50)	(0.50)	(0.00)
Native	0.80	0.83	0.77	0.83	-0.03***
	(0.40)	(0.37)	(0.42)	(0.38)	(0.00)
Labour-market[b]					
Employed	0.90	0.95	0.83	0.95	-0.06***
	(0.30)	(0.21)	(0.37)	(0.22)	(0.00)
Gross earnings (2010 euros)	22,461	30,894	20,778	30,553	-1,334.4***
	(22,319)	(26,963)	(18,263)	(26,035)	(116.9)
UI recipient	0.11	0.03	0.18	0.05	0.05***
	(0.32)	(0.18)	(0.38)	(0.22)	(0.00)
Welfare recipient	0.03	0.02	0.04	0.02	0.01***
	(0.18)	(0.13)	(0.19)	(0.13)	(0.00)
Health status					
Health index	0.18	0.03	0.27	0.04	0.09***
	(0.86)	(0.32)	(1.13)	(0.39)	(0.002)
Hosp. in the past three years[c]	0.35	0.16	0.40	0.16	0.05***
	(0.47)	(0.36)	(0.49)	(0.37)	(0.00)
Dead within five years	0.03	0.01	0.04	0.01	0.01***
	(0.17)	(0.10)	(0.20)	(0.10)	(0.00)
NSII final decision					
Award rate (cond.	0.60	.	0.62	.	.
on applying)	(0.49)	.	(0.49)	.	.
Number of observations	158,390	1,239,579	81,912	1,271,414	2,751,295

Notes : [a] To compute these summary statistics we take – in each year – the full sample of applicants (that year) as well as a 10% random sample of non-applicants (that year). [b] Our sample excludes individuals not employed in the previous year (see Section 4). [c] Hospitalized in t-1; t-2 or t-3. Column (5) presents the difference-in-differences mean, i.e. for each variable Y $(Y_{applicant,after} - Y_{applicant,before}) - (Y_{non-applicant,after} - Y_{non-applicant,before})$.

Table 3: Death rate within five years by screening decisions (%).

SCREENING STAGE	TOTAL	BY DIRECTION OF OUTCOME	
		Positive (application / allowance)	Negative (no application / denial)
Panel A : Before Gatekeeper (2001-2002)			
Application (direct/self-screening)	1.23	3.05	0.99
	(1,397,969)	(158,390)	(1,239,579)
Award decision	3.05	4.18	1.45
	(158,390)	(94,945)	(38,373)
Panel B : After Gatekeeper (2003-2004)			
Application (direct/self-screening)	1.18	4.14	0.99
	(1,353,326)	(81,912)	(1,271,414)
Award decision	4.14	5.67	1.50
	(81,912)	(50,550)	(23,830)

Notes : Individuals may apply to DI and recover before the final examination, so the total number of applicants is larger than the sum of allowed and denied applicants.

Table 4: Summary statistics of the data (10% random sample)[a]

	2001-2002		2003-2004		Diff-in-
	Treated	Control	Treated	Control	diff
	(1)	(2)	(3)	(4)	(5)
Demographics					
Age	41.17	40.95	41.81	41.77	-0.015
	(10.10)	(41.11)	(10.22)	(10.20)	(0.05)
Male	0.59	0.57	0.58	0.56	-0.003
	(0.49)	(0.50)	(0.49)	(0.50)	(0.002)
Native	0.87	0.83	0.87	0.83	0.005**
	(0.33)	(0.37)	(0.33)	(0.38)	(0.002)
Labour-market[b]					
Employed	0.95	0.95	0.95	0.95	0.002
	(0.21)	(0.21)	(0.21)	(0.22)	(0.001)
Gross earnings (2010 euros)	28,896	30,987	28,190	30,142	112.9
	(22,054)	(27,141)	(21,420)	(25,468)	(127.7)
UI recipient	0.03	0.04	0.05	0.05	0.000
	(0.18)	(0.18)	(0.22)	(0.23)	(0.001)
Welfare recipient	0.01	0.01	0.01	0.02	0.001
	(0.12)	(0.13)	(0.12)	(0.13)	(0.001)
DI					
DI application rate	0.014	0.013	0.007	0.006	-0.001
	(0.12)	(0.11)	(0.08)	(0.08)	(0.000)
DI award rate	0.008	0.008	0.004	0.004	-0.000
	(0.09)	(0.09)	(0.07)	(0.06)	(0.000)
Health status					
Health index	0.034	0.034	0.043	0.042	0.000
	(0.34)	(0.34)	(0.39)	(0.40)	(0.002)
Hospitalized in the three previous years[c]	0.16	0.16	0.17	0.17	-0.001
	(0.37)	(0.37)	(0.38)	(0.37)	(0.002)
Hospitalization type (among hospitalized)[d]					
Musculo-skeletal disorders	0.20	0.19	0.19	0.18	0.010*
	(0.40)	(0.40)	(0.40)	(0.39)	(0.005)
Neoplams	0.06	0.06	0.06	0.063	-0.002
	(0.23)	(0.24)	(0.23)	(0.24)	(0.003)
Cardiovascular diseases	0.10	0.10	0.10	0.10	-0.002
	(0.30)	(0.30)	(0.30)	(0.30)	(0.004)
Mental disorders	0.01	0.01	0.01	0.01	0.002*
	(0.11)	(0.10)	(0.12)	(0.10)	(0.001)
Endocrine problems	0.01	0.01	0.01	0.01	0.002
	(0.11)	(0.12)	(0.12)	(0.12)	(0.001)
Nervous disorders	0.06	0.06	0.07	0.07	0.002
	(0.24)	(0.24)	(0.26)	(0.25)	(0.003)
Dead within five years	0.01	0.01	0.01	0.01	0.000
	(0.10)	(0.10)	(0.10)	(0.01)	(0.000)
Number of observations	89,389	1,146,014	90,984	1,166,281	2,492,668

Notes : [a] To compute these summary statistics we take a 10% random sample of applicants and non-applicants in each year (2001-2004). [b] Our sample excludes individuals not employed in the previous year (see Section 4). [c] Hospitalized in t-1; t-2 or t-3. [d] Note that only the main hospitalization types are listed here. Individuals can be hospitalized for several reasons in the previous three years, so lines could add up to more than 100%. Column (5) presents the difference-in-differences mean, i.e. for each variable Y $(Y_{treated,after} - Y_{treated,before}) - (Y_{control,after} - Y_{control,before})$.

Table 5: Summary statistics for the full sample of applicants[a]

	2001-2002		2003-2004		Diff-in-
	Treated	Control	Treated	Control	diff
	(1)	(2)	(3)	(4)	(5)
Health status of DI applicants					
Health index	0.163	0.179	0.252	0.275	-0.006
	(0.794)	(0.884)	(1.033)	(1.143)	(0.016)
Hospitalized in the three previous years	0.336	0.343	0.405	0.396	0.015*
	(0.472)	(0.475)	(0.491)	(0.489)	(0.008)
Dead within five years	0.028	0.031	0.036	0.042	-0.003
	(0.166)	(0.173)	(0.187)	(0.200)	(0.003)
Type of impairment of DI applicants					
Musculo-skeletal disorders	0.290	0.288	0.303	0.285	0.016*
	(0.454)	(0.453)	(0.460)	(0.451)	(0.007)
Mental disorders	0.401	0.365	0.278	0.287	-0.045***
	(0.490)	(0.481)	(0.448)	(0.453)	(0.008)
Cardiovascular diseases	0.047	0.048	0.061	0.059	0.002
	(0.212)	(0.213)	(0.239)	(0.236)	(0.004)
Nervous disorders	0.035	0.037	0.047	0.046	0.003
	(0.183)	(0.189)	(0.212)	(0.211)	(0.003)
Respiratory disorders	0.014	0.015	0.016	0.018	0.000
	(0.116)	(0.123)	(0.127)	(0.132)	(0.002)
Endocrine problems	0.010	0.011	0.012	0.014	0.000
	(0.099)	(0.106)	(0.108)	(0.116)	(0.002)
Other	0.203	0.235	0.283	0.290	0.017**
	(0.402)	(0.424)	(0.450)	(0.454)	(0.007)
NSII Decision					
DI award rate for applicants					
All impairment types	0.609	0.599	0.632	0.616	0.005
	(0.488)	(0.490)	(0.482)	(0.486)	(0.008)
Musculo-skeletal disorders	0.727	0.676	0.584	0.581	-0.047***
	(0.446)	(0.468)	(0.493)	(0.493)	(0.014)
Mental disorders	0.538	0.592	0.705	0.676	0.084***
	(0.499)	(0.491)	(0.456)	(0.468)	(0.014)
Cardiovascular diseases	0.864	0.818	0.791	0.776	-0.030
	(0.343)	(0.386)	(0.407)	(0.417)	(0.027)
Nervous disorders	0.829	0.826	0.795	0.807	-0.016
	(0.377)	(0.379)	(0.404)	(0.394)	(0.030)
Respiratory disorders	0.802	0.792	0.784	0.769	0.005
	(0.399)	(0.406)	(0.413)	(0.422)	(0.054)
Endocrine problems	0.852	0.775	0.743	0.727	-0.061
	(0.356)	(0.417)	(0.440)	(0.446)	(0.065)
Other	0.459	0.412	0.537	0.514	-0.025
	(0.498)	(0.492)	(0.499)	(0.500)	(0.016)
Degree of disability (among awarded)					
Fully disabled (disability degree > 80%)	0.539	0.516	0.565	0.531	0.011
	(0.499)	(0.500)	(0.496)	(0.499)	(0.010)
Number of observations	12,264	144,505	6,231	74,676	237,676

Notes : [a] To compute these summary statistics, we take all applicants each year between 2001 and 2004. Column (5) presents the difference-in-differences mean, i.e. for each variable Y $(Y_{treated,after} - Y_{treated,before}) - (Y_{control,after} - Y_{control,before})$.

Table 6: The effect of stricter screening on DI application rate : difference-in-differences estimates

DI application rate	
	Model (1)
Treatment (stricter screening)	-0.005* [-4.4%]
	(0.003)
Year dummies	✓
Regional dummies	✓
Nb. of obs.	2,706,577

Notes : (1) Standard errors are clustered at the regional level (2) Estimates in brackets are presented in percentage change of the baseline DI application rate (0.1132).

Table 7: The effect of stricter screening on DI application rate : difference-in-differences estimates; interacting treatment with year dummies.

	DI application rate		
	in 2001	in 2003	in 2004
Treatment (stricter screening)	0.002 [+1.9%]	-0.006** [-5.8%]	-0.002 [-1.9%]
	(0.002)	(0.002)	(0.003)
Nb. of obs.		2,706,577	

Notes : (1) Our model includes year and regional dummies as well as a series of interactions terms treatment*year dummies. We leave out the one interaction term for the last pre-treatment period (2002) – see Equation 10 (2) Standard errors are clustered at the regional level (3) Estimates in brackets are presented in percentage change of the baseline DI application rate (0.1132).

Table 8: The effect of stricter screening : coefficients from individual OLS estimates

	DI application rate	
	coeff	s.e
Treatment (stricter screening)	-0.005*	(0.003)
Male	-0.045***	(0.001)
Dutch	-0.032***	(0.002)
Age (ref. category :[25-29])		
Age [30-34]	0.011***	(0.001)
Age [35-39]	0.019***	(0.002)
Age [40-44]	0.029***	(0.002)
Age [45-49]	0.041***	(0.002)
Age [50-54]	0.059***	(0.002)
Age [55-59]	0.070***	(0.002)
Age [60-65]	0.022***	(0.002)
Year dummies	✓	
Regional dummies	✓	
Nb. of obs.	2,706,577	

Notes : (1) Standard errors are clustered at the regional level.

Table 9: The effect of stricter screening – Heterogeneity by past health status

	DI application rate		
	2003	2004	N
By health index :			
Health index=0	-0.005** [-4.7%]	-0.002 [-1.9%]	*2,637,761*
	(0.002)	(0.003)	
Health index=1	-0.026 [-8.1%]	0.006 [-1.9%]	*34,265*
	(0.025)	(0.013)	
Health index $\geq$ 2	-0.013 [-3.1%]	-0.008 [-1.9%]	*34,551*
	(0.017)	(0.009)	
By hospitalization in the three previous years :			
Not hospitalized	-0.006*** [-6.6%]	-0.004$^{\mu}$ [4.4%]	*2,222,146*
	(0.002)	(0.002)	
Hospitalized (all-cause)	-0.003 [-1.4%]	0.005 [+2.3%]	*484,431*
	(0.006)	(0.006)	
Hospitalized – cause-specific			
Musculo-skeletal	-0.018* [-7.7%]	-0.000 [-0.1%]	*92,738*
	(0.010)	(0.017)	
Neoplasms	-0.006 [-1.6%]	-0.011 [-3.0%]	*34,806*
	(0.025)	(0.008)	
Cardiovascular	-0.027*** [-8.4%]	0.204 [+63.6%]	*52,231*
	(0.008)	(0.013)	
Mental	-0.034$^{\mu}$ [-8.6%]	0.037 [+9.4%]	*5,541*
	(0.024)	(0.059)	
Endocrine	-0.035 [-14.2%]	0.024 [+9.8%]	*7,016*
	(0.025)	(0.069)	
Nervous	-0.023*** [-9.5%]	0.005 [+2.1%]	*32,037*
	(0.006)	(0.008)	
Year dummies	✓	✓	
Regional dummies	✓	✓	
Background individual characteristics	✓	✓	

Notes : (1) Standard errors are clustered at the regional level. (2) Each line presents the estimated coefficient associated with the treatment (stricter screening) for a sub-sample of individuals defined by their previous health status. Estimates in brackets are presented in percentage change of the baseline DI application rate for the sub-sample of interest.

Table 10: The effect of stricter screening on impairment-specific DI application.

	Treatment (stricter screening)	
	2003	2004
DI application for impairment type :		
Musculo-skeletal	-0.001 [-3.1%]	-0.001 [-3.1%]
	(0.001)	(0.001)
Mental disorders	-0.006** [-14.5%]	-0.005** [-12.1%]
	(0.003)	(0.002)
Cardiovascular diseases	-0.000 [-5.5%]	-0.000 [+11.1%]
	(0.000)	(0.000)
Nervous disorders	-0.000 [-0.1%]	-0.001 [-9.5%]
	(0.000)	(0.000)
Respiratory disorders	-0.000 [-5.8%]	-0.000 [+5.8.0%]
	(0.000)	(0.000)
Endocrine problems	-0.000 [-15.4%]	0.000* [+7.7%]
	(0.000)	(0.000)
Other	0.002 [+7.5%]	0.004** [+15.0%]
	(0.001)	(0.002)
Year dummies	✓	✓
Regional dummies	✓	✓
Background individual characteristics	✓	✓
Nb. of obs.	2,706,577	

Notes : (1) Standard errors are clustered at the regional level. (2) Each column presents the estimated coefficient associated with the treatment (stricter screening) for a different outcome : DI application for a specific impairment type. Estimates in brackets are presented in percentage change of the baseline impairment-specific DI application rate.

Table 11: The effect of stricter screening on the composition of the pool of DI applicants (on the year of application).

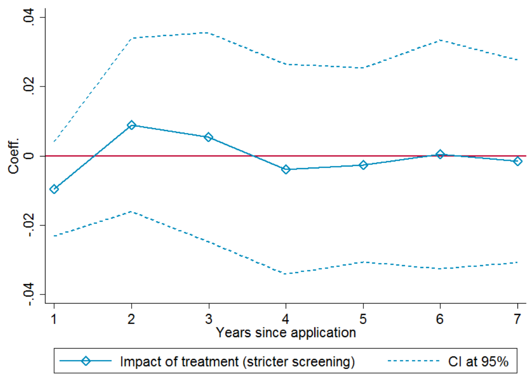
	Treatment (stricter screening)	
	2003	2004
Health status		
Health index	0.003 [+1.7%]	0.007 [+3.9%]
	(0.015)	(0.020)
Hospitalized in the three previous years	0.013*** [+3.8%]	0.016** [+4.7%]
	(0.005)	(0.007)
Share of impairments (as defined by NSII)		
Musculo-skeletal disorders	0.017* [+5.9%]	0.006 [+2.1%]
	(0.010)	(0.007)
Mental disorders	-0.038** [-10.4%]	-0.040*** [-11.0%]
	(0.017)	(0.010)
Cardiovascular diseases	-0.000 [-0.6%]	0.002 [+4.2%]
	(0.002)	(0.003)
Nervous disorders	0.002 [+5.4%]	0.001 [+2.7%]
	(0.006)	(0.004)
Respiratory disorders	0.000 [+3.2%]	0.001 [+6.5%]
	(0.002)	(0.003)
Endocrine problems	-0.003 [-26.0%]	0.001 [+8.8%]
	(0.002)	(0.003)
Other	0.022** [+9.4%]	0.029** [+12.4%]
	(0.008)	(0.011)
Share receiving full DI	0.014 [+2.1%]	0.013 [+2.5%]
	(0.030)	(0.008)
Year dummies	✓	✓
Regional dummies	✓	✓
Background individual characteristics	✓	✓
Nb. of obs.	317,798	

Notes : (1) Standard errors are clustered at the regional level. (2) Each line presents the estimated coefficient associated with the treatment (stricter screening) for a given health variable among the sample of DI applicants. Estimates in brackets are presented in percentage change of the baseline health status among DI applicants.

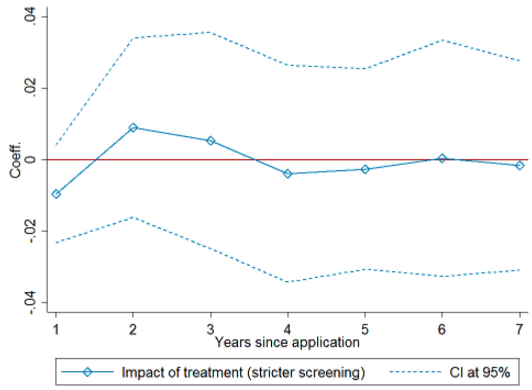
A Figures

Figure 5: The short and long-run effect of stricter screening on labour and health outcomes for DI applicants.

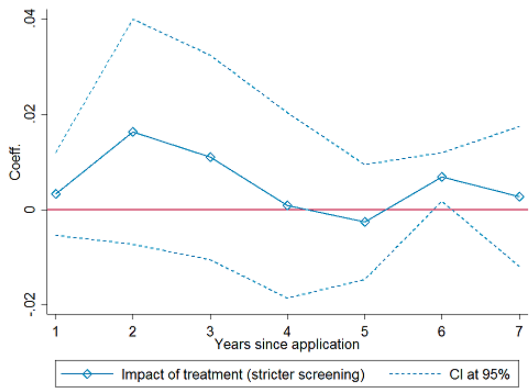
(a) Earnings



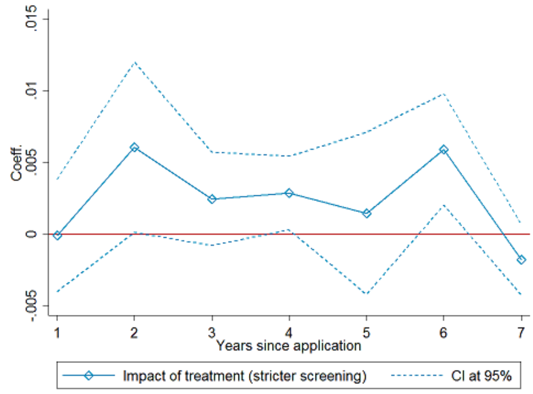
(b) Employment



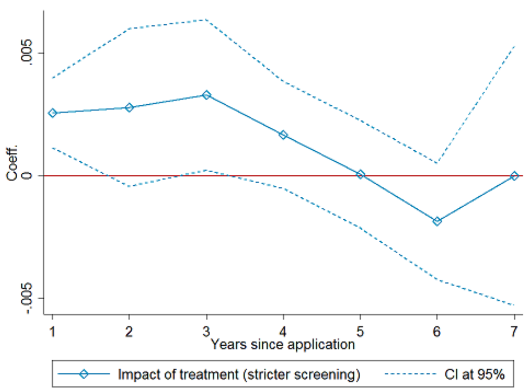
(c) UI receipt



(d) Welfare receipt

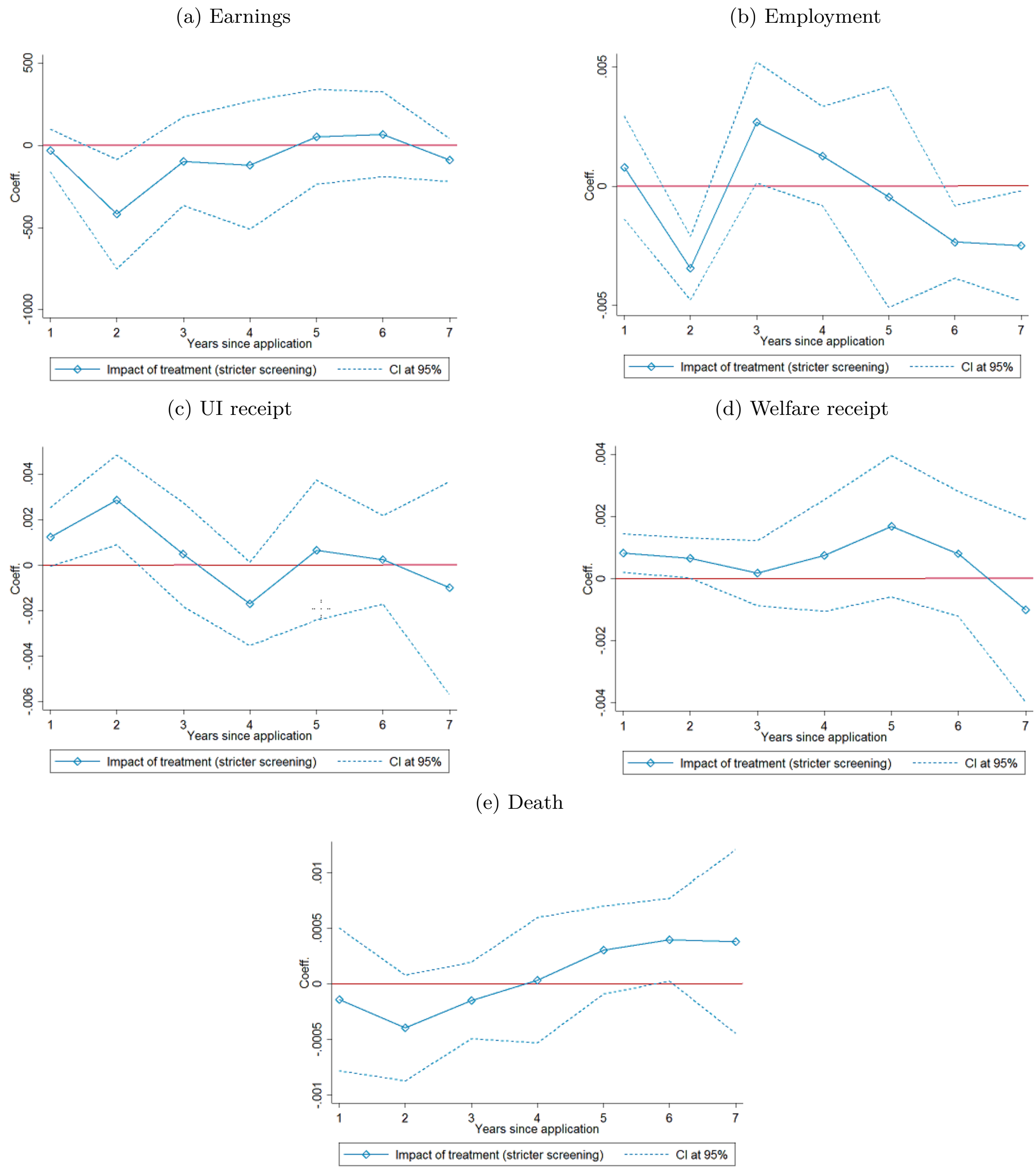


(e) Death



Note : Treatment effects for each year since (potential) DI application are computed by estimating Equation (10). We show the coefficient associated with year 2003.

Figure 6: The short and long-run effect of stricter screening on labour and health outcomes for DI non-applicants.



Note : Treatment effects for each year since (potential) DI application are computed by estimating Equation (10). We show the coefficient associated with year 2003.

B Tables

Table A1: Earnings of disabled-worker determinations by screening decisions (%).

SCREENING STAGE	TOTAL	BY DIRECTION OF OUTCOME	
		Positive (application or allowance)	Negative (non application or denial)
Panel A : Before Gatekeeper (2001-2002)			
Pre-application (direct of self-screening)	29,938	22,461	30,893
	(1,397,969)	(158,390)	(1,239,579)
Active screening (NSII final examination)	22,461	22,893	21,695
	(158,390)	(94,945)	(38,373)
Panel B : After Gatekeeper (2003-2004)			
Pre-application (direct of self-screening)	29,961	20,778	30,553
	(1,353,326)	(81,912)	(1,271,414)
Active screening (NSII final examination)	20,778	21,262	18,231
	(81,912)	(50,550)	(23,830)

Notes : Individuals may apply to DI and recover before the final examination, so the total number of applicants is larger than the sum of allowed and denied applicants. Earnings are measured on the year of (potential) application.

Table A2: Employment rate of disabled-worker determinations by screening decisions (%).

SCREENING STAGE	TOTAL	BY DIRECTION OF OUTCOME	
		Positive (application or allowance)	Negative (non application or denial)
Panel A : Before Gatekeeper (2001-2002)			
Pre-application (direct of self-screening)	94.6	89.8	95.2
	(1,397,969)	(158,390)	(1,239,579)
Active screening (NSII final examination)	89.8	87.7	90.4
	(158,390)	(94,945)	(38,373)
Panel B : After Gatekeeper (2003-2004)			
Pre-application (direct of self-screening)	94.1	83.2	94.8
	(1,353,326)	(81,912)	(1,271,414)
Active screening (NSII final examination)	83.2	82.1	83.2
	(81,912)	(50,550)	(23,830)

Notes : Individuals may apply to DI and recover before the final examination, so the total number of applicants is larger than the sum of allowed and denied applicants. Employment is measured on the year of (potential) application.

Table A3: Unemployment rate of disabled-worker determinations by screening decisions (%).

SCREENING STAGE	TOTAL	BY DIRECTION OF OUTCOME	
		Positive (application or allowance)	Negative (non application or denial)
Panel A : Before Gatekeeper (2001-2002)			
Pre-application (direct of self-screening)	4.2	11.2	3.3
	(1,397,969)	(158,390)	(1,239,579)
Active screening (NSII final examination)	11.2	9.5	20.5
	(158,390)	(94,945)	(38,373)
Panel B : After Gatekeeper (2003-2004)			
Pre-application (direct of self-screening)	5.9	17.7	5.2
	(1,353,326)	(81,912)	(1,271,414)
Active screening (NSII final examination)	17.7	12.0	30.9
	(81,912)	(50,550)	(23,830)

Notes : Individuals may apply to DI and recover before the final examination, so the total number of applicants is larger than the sum of allowed and denied applicants. Unemployment is measured on the year of (potential) application.

Table A4: Welfare receipt rate of disabled-worker determinations by screening decisions (%).

SCREENING STAGE	TOTAL	BY DIRECTION OF OUTCOME	
		Positive (application or allowance)	Negative (non application or denial)
Panel A : Before Gatekeeper (2001-2002)			
Pre-application (direct of self-screening)	1.8	3.2	1.6
	(1,397,969)	(158,390)	(1,239,579)
Active screening (NSII final examination)	3.2	2.0	6.2
	(158,390)	(94,945)	(38,373)
Panel B : After Gatekeeper (2003-2004)			
Pre-application (direct of self-screening)	1.7	3.7	1.6
	(1,353,326)	(81,912)	(1,271,414)
Active screening (NSII final examination)	3.7	2.0	6.6
	(81,912)	(50,550)	(23,830)

Notes : Individuals may apply to DI and recover before the final examination, so the total number of applicants is larger than the sum of allowed and denied applicants. Welfare receipt is measured on the year of (potential) application.

C Data appendix : The Charlson Comorbidity Index

The Charlson Comorbidity Index (CCI) is a popular tool for predicting mortality by classifying or weighting comorbid conditions (comorbidities) (Charlson et al., 1987). More specifically, it reflects the cumulative increase in likelihood of one-year mortality due to the severity of the effect of comorbidities. The CCI can be constructed from medical record abstract or administrative data.

We use the coding algorithm developed by Stagg et al. (2015) to derive the CCI from ICD-9-CM administrative data. Our index is a weighted sum of 17 comorbidities. The 17 comorbidities and their associated weights – that allow for adjustment for severity of illness – are listed in Table A1.

Table A1: Charlson Comorbidities and Weights

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