SCREENING DISABILITY INSURANCE APPLICATIONS

Philip de Jong
University of Amsterdam
and Aarts, De Jong, Wilms &
Goudriaan Public Economics B.V.

Maarten Lindeboom
VU University Amsterdam
Tinbergen Institute and Health
Economics Bergen

Bas van der Klaauw
VU University Amsterdam
Tinbergen Institute

Abstract
This paper investigates the effects of stricter screening of disability insurance applications. A large-scale experiment was set up wherein two of the 26 Dutch regions caseworkers of the disability insurance administration were instructed to screen applications more stringently. The empirical results show that stricter screening reduces long-term sickness absenteeism and disability insurance applications. We find evidence for direct effects of stricter screening on work resumption during the period of sickness absence and for self-screening by potential disability insurance applicants. Furthermore, stricter screening improves targeting efficiency without inducing negative spillover effects on the inflow into unemployment insurance. (JEL: C93, H53, I18)

1. Introduction

Social insurance programs covering for example sickness absenteeism, unemployment, or disability are always and everywhere subject to moral hazard. The degree of moral hazard depends on the accessibility and generosity of the program. Various policy measures are in place to reduce the attractiveness of social insurance programs. Examples are imposing punitive monetary sanctions on unemployment benefit recipients for voluntary job loss, and reduced or no sick pay during the first few days of absenteeism. An alternative to such financial instruments is to increase the level of

This paper was edited by Orazio Attanasio
Acknowledgments: The authors thank Jonneke Bolhaar for excellent research assistance. We furthermore would like to thank Orazio Attanasio, John Bound, Norma Coe, Pilar Garcia Gomez, Laura Hartman, three anonymous referees, seminar participants at Uppsala University, Utrecht University, Wuerzburg University, Nuremberg University, CREST-Paris, and Universidade Técnica de Lisboa, and participants of the 15th workshop in Health Econometrics in Thessaloniki, the 2006 European association of Labor Economics (EALE) conference in Prague, and the 2006 Econometric Society Meeting (ESEM) in Vienna for useful comments and suggestions. Lindeboom is a member of the Netspar and van der Klaauw is a member of the CEPR.

E-mail: ph.dejong@ape.nl (de Jong); mlindeboom@feweb.vu.nl (Lindeboom); bklaauw@feweb.vu.nl (van der Klaauw)

Journal of the European Economic Association February 2011 9(1):106–129

DOI: 10.1111/j.1542-4774.2010.01002.x

screening. Active screening policies can be useful in reducing information asymmetries and hence improve the eligibility determination process in social insurance programs. Screening of claims is common in many insurance markets, such as indemnity insurance and fire insurance. Furthermore, screening of applications is an important aspect of the provision of social care and long-term health care for the elderly. In this paper we investigate to what extent the level of screening can be used to control the inflow into disability insurance (DI).

In most OECD countries DI programs are substantial in size and have experienced strong growth in recent decades (OECD 2003). With little screening of DI applications the program may be used by workers to leave the labor force, and by employers to lay off redundant workers. Screening of DI applications therefore plays an important role in the application procedure of DI programs around the world.[1] We focus on the effects of stricter screening on the incidence of long-term sickness absence and DI benefit applications in the Netherlands. The Dutch DI program was notorious for its excessive growth in the 1970s and 1980s (see Bound and Burkhauser 1999; Burkhauser, Daly, and De Jong 2008). Since then, a series of reforms have been implemented to reduce the fiscal burden of the DI program. One of the most recent reforms was the introduction of a different screening procedure in the application process.

Stricter screening of benefit claims may overcome moral hazard problems. It forces employers to devote serious effort to getting sick workers back to work before they enter the DI program. If reintegration activities are effective, increased effort positively affects work resumption during sickness absenteeism. Furthermore, stricter screening of reintegration effort reduces the attractiveness of the DI program to potential applicants. This may trigger a mechanism of self-selection or self-screening (Halpern and Hausman 1986; Parsons 1991).

Previous research in this area mostly relied on state level variation in the implementation of DI rules (e.g. Autor and Duggan 2003; Gruber 2000; Gruber and Kubik 1997; Parsons 1991). 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. If the latter is the case, high denial rates may merely reflect loose screening practices in the past, rather than a strict screening policy.

The data used in this paper are from a controlled experiment in which regional variation in screening intensity is determined exogenously. 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

1. See Bound and Burkhauser (1999) also for a review of studies on the disincentive effects of DI programs. See Chen and Van der Klaauw (2008) for a recent study on the disincentive effects of the DI program in the Unites States and see Johansson and Palme (2002) for a recent contribution on the effect of the level of sick pay on long-term absence in Sweden.

than elsewhere. The behavior of caseworkers was monitored and the screening practice in treatment regions was indeed found to be stricter than in other regions. In treatment regions time spent on screening reintegration reports by caseworkers was 40% higher than in control regions. The difference in screening practice between treatment and control regions was not announced beforehand. In the empirical analyses we use data from administrative records of the NSII; we use both difference-in-difference analyses at the level of regions and estimate more parameterized Logit models at the individual level. We find that stricter screening decreases long-term sickness absenteeism (defined as spells lasting at least 13 weeks) and DI applications equally. We argue that the reduction in long-term sickness absenteeism is due to self-screening by potential DI applicants and that for DI applications the decline is due to a direct effect on work resumption during sickness absence.

An important issue in this literature is targeting efficiency (Gruber and Kubik 1997; Parsons 1991). Parsons shows that under certain conditions increases in the DI benefit denial rate can lead to perverse effects, leading disabled workers to reduce applications more than able workers. From the institutional background in the Netherlands and some sensitivity analyses, we argue that targeting of the Dutch DI system is likely to have improved and that there are no perverse effects of stricter screening.

The outline of this paper is as follows. In Section 2 we provide some institutional background concerning the Dutch DI system. Section 3 discusses the setup of the experiment. Section 4 gives some theory. In Section 5 we discuss the administrative database that will be used in the empirical analyses. In Section 6 we present our estimation results. Section 7 concludes.

2. Institutional Background

In this section we describe the Dutch DI system as it was in 2003, the year the experiment took place. Furthermore, we discuss the reintegration activities to which sick workers and their employers have to comply. We only discuss aspects that are relevant for this study.

2.1. Sickness and DI Benefits

The Dutch DI program covers all employees. Any illness or injury triggers entitlement within the DI program after a mandatory waiting period of one year. The legitimacy of sickness absenteeism during the waiting period is checked by a doctor from an occupational health service contracted by the employer. During the one year waiting period employers are responsible for financing sick pay.[2] The employers should furthermore contract an occupational health service to prevent and manage sickness absenteeism.

2. Collective bargaining agreements ensure that sick workers receive 90–100% of their net salary. Most small- and medium-sized employers insure themselves against financing sick pay.

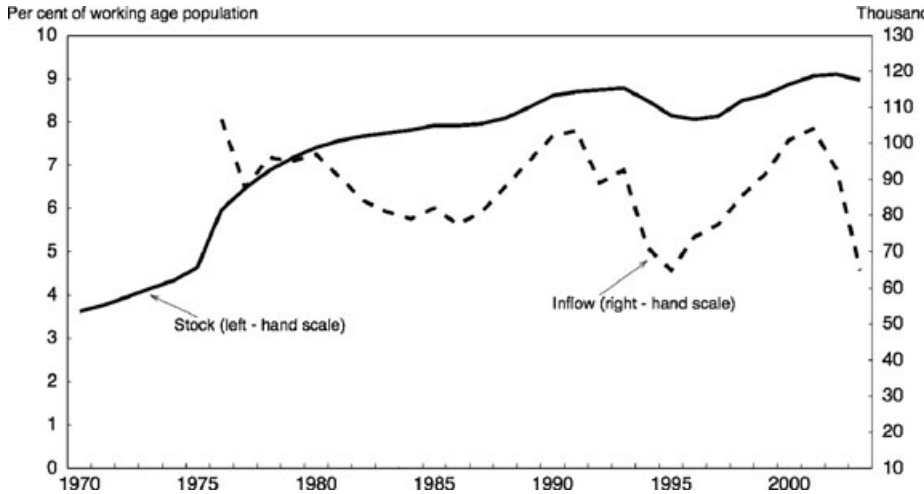

FIGURE 1. DI benefit recipients: stock and inflow 1970–2003.
Source: OECD (2004)

While other OECD countries make a distinction according to whether the impairment occurred on the job or elsewhere, only the consequence of the impairment is relevant for the Dutch DI program. The degree of disability is assessed by considering a worker's residual earning capacity, defined by potential earnings with the worker's functional limitations as a fraction of pre-disability earnings. The degree of disability is the complement of the residual earning capacity, and is measured in seven classes, ranging from a 15% to a 100% loss. Both disability and unemployment insurance (UI) schemes are administered by the NSII. When considered fully disabled by the NSII, 70% of pre-disability earnings are replaced. If assessed as partially (15%–80%) disabled the replacement rate is correspondingly lower. Benefits are capped at about € 32,000. The size of DI benefits and its unlimited duration (until pension age 65) make DI a more attractive option than UI. UI benefits have a limited entitlement period, after which the unemployed can apply for means-tested welfare benefits.[3]

Figure 1 shows the number of individuals collecting DI benefits as percentage of the working age population and the number of inflows into DI. The figure shows a sharp increase in the percentage of individuals collecting DI benefits in the 1970s and 1980s. The number of beneficiaries increased from 475,000 in 1976 to 921,000 in 1993. During the 1990s the disability definition was narrowed and benefit generosity was reduced. This lowered the inflow into DI. Furthermore, between 1994 and 1996 part

3. The DI benefit period is cut into two chronologically linked parts. The first is a short-term wage-related benefit replacing 70% of before-tax earnings, when assessed as fully disabled. The duration of this wage-related benefit depends on age at the onset of disability. It varies from zero for those under age 33 to six years for those whose disability started at age 58 or beyond. The second part is a so-called follow-up benefit with a lower income base. This earnings base is the minimum wage plus a supplement depending on age at onset according to the formula: $0.02 \times$ [age at onset $-$ 15] $\times$ [wage $-$ minimum wage]. Most collective bargaining agreements covered the gap between the lower benefits in the follow-up period and the wage-related benefits in the initial period of disability.

of the stock of beneficiaries was re-examined using the new, more stringent, eligibility rules. As a result, the number of benefit terminations grew sharply and, on balance, led to a 7% drop in the number of beneficiaries to 855,000 in 1996. From then on, the numbers started growing again and reached 979,000 in November 2002.

2.2. The Gatekeeper Protocol and the Role of Reintegration Reports

Until April 2002 the NSII had during the one year waiting period of sickness absenteeism a joint responsibility with the sick worker and the employer to get the sick worker back to work. As of April 2002, the so-called gatekeeper protocol was introduced, which implies that the NSII is no longer involved in the process of reintegrating sick workers during the waiting period, but purely acts as a gatekeeper. This protocol stipulates the responsibilities of the worker and the employer for sick spells lasting six weeks or longer. Figure 2 gives a schematic representation of the different steps taken in this protocol. After a maximum of six weeks of absence the doctor from the occupational health service has to make a first assessment of medical cause and functional limitations, and gives a prognosis regarding work resumption. 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 current/other job under current/accommodated conditions) and the steps needed to reach that aim. They appoint a case manager, and fix dates at which the plan should be evaluated and modified if necessary. The 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

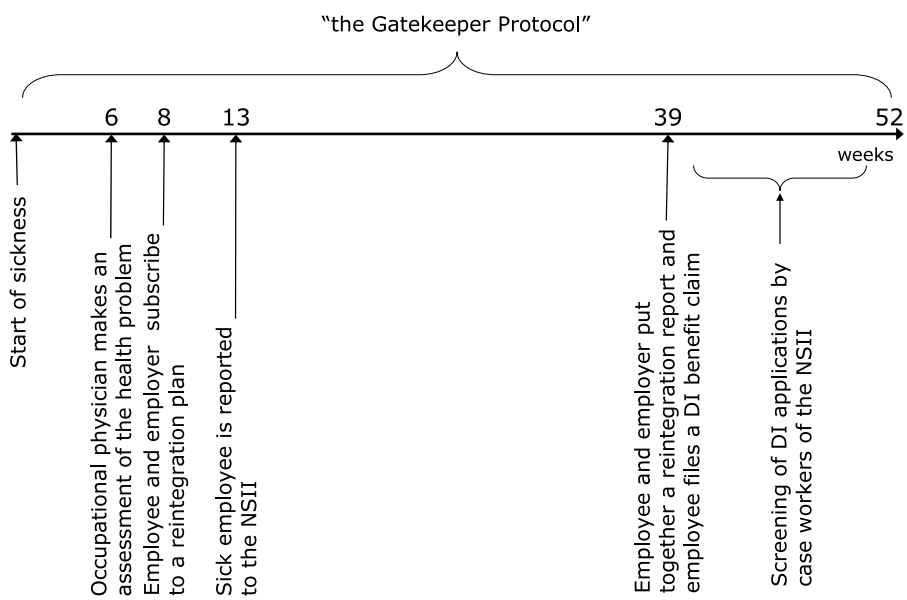

FIGURE 2. Schematic representation of the process toward entering DI.

report the sick employee to the NSII, which is only a paper obligation. From this moment on the worker is added to the administrative database of the NSII.

If the worker has not fully returned to work before 39 weeks of sickness, he or she files a DI benefit claim and starts collecting benefits after twelve months, if the application is approved. Benefit claims are only considered admissible if they are accompanied by a reintegration report, containing the original reintegration plan, and an assessment as to why the plan has not (yet) resulted in work resumption. A caseworker of the NSII screens the reintegration report. The claim is not processed if the report is delayed, incomplete, or if it is clear that reintegration efforts were insufficient. Depending on the seriousness of the negligence the caseworker can return the reintegration report and give the employer the opportunity to complete it or the caseworker can start a sanction procedure against the employer. A sanction usually implies that the employer is obliged to continue providing sick pay for some additional months.

The employer and the sick employee 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. Part of the employer's responsibility, therefore, is to penalize an unwilling employee. All workers becoming sick after April 1, 2002 were treated according to the gatekeeper protocol. Since DI applications are made in the 39th week of sickness absence the gatekeeper protocol started to affect DI claims as of January 2003. Note that reintegration reports and the screening of these were a new element in the DI application procedure. Hence, NSII caseworkers, employers, workers, and the occupational health services had no experience with the screening of reports prior to January 2003.

3. The Experiment

Our experiment started January 1, 2003, when the first DI applications under the new gatekeeper protocol arrived at the regional offices of the NSII. As already mentioned, these were the first DI applications that were accompanied by reintegration reports. The standard (national) procedure was to screen reintegration reports "on paper" and to only contact the employer and/or the sick worker directly if there was some suspicion of negligence and/or fraud. In two out of the 26 regions we instructed the caseworkers to implement a screening policy that was stricter than elsewhere.

In the treatment regions caseworkers were instructed to always contact or visit the employer and/or the sick employee, unless it was absolutely clear that, given the medical condition, sufficient efforts were already being undertaken.[4] Stricter screening

4. The main reason to screen a reintegration report only "on paper" in treatment regions is that the caseworker already handled multiple applicants from an employer and that therefore the caseworker is familiar with the reintegration protocol of this employer.

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 occupational health agency	3%	12%	32%
Visit to employer	9%	41%	7%
Face to face contact with worker	77%	41%	18%
Unknown	4%	2%	

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

was only applied to individuals who had not resumed work for more than 50% of their contractual working hours and were not severely and permanently disabled. The experiment period finished at the end of October 2003. We set up the experiment together with the NSII, which made additional resources available for the treatment regions to intensify the screening. The experiment was not announced beforehand to any of the parties involved (the worker, the employer, and the occupational health services). We designed the experiment, but the treatment regions were assigned to us by the NSII. In Section 5, we use pre-experiment outcomes on sickness absenteeism (more precisely, 13-week sickness reports to the NSII), DI applications at week 39, and UI inflows to investigate if treatment assignment was actually independent of these outcomes.

Because the start of the experiment coincided with the moment the gatekeeper protocol became effective, none of the caseworkers at the NSII, independent of their location, had any experience in screening reintegration reports. For the success of the experiment it was crucial that in the treatment regions the new screening regime was conducted more strictly than in the rest of the Netherlands. To monitor the screening regimes in all regions we sent out questionnaires to the 24 control regions. In the treatment regions we required caseworkers to fill out a questionnaire for each DI application. In the control regions we asked the local managers to regularly report summaries of the screening regime at their offices. Table 1 summarizes the difference in screening intensity in the treatment and the control regions. The table shows that in the treatment regions caseworkers visited sick workers (Apeldoorn region) or their employers (Hengelo region) more often than elsewhere. We return to the difference in effectiveness of both screening methods. 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.

4. Potential Effects of Screening

Employers are financially responsible for sick pay during the first 12 months of absence. The employer drafts together with the sick worker a reintegration plan, but usually it is

the employer who decides on reintegration activities and workplace accommodations needed to affect work resumption. Recall from Section 2.2 that the NSII holds the employer responsible in case of negligence. We therefore take the employer as the relevant decision-making agent.

The introduction of the gatekeeper protocol imposes minimum requirements on reintegration efforts that the employer should offer, and the employer is at risk of getting a sanction if the actual reintegration efforts do not meet these minimum requirements. The probability of getting a sanction and the size of the sanction increase with the extent of noncompliance. Reintegration efforts are costly to employers, at a minimum because they are mandated to contract an occupational health agency to manage absenteeism. The occupational health agencies usually work for many different employers and thus collect and transmit information about the screening policy at the local offices of the NSII. We return to this issue in what follows.

Employers choose their reintegration activities such that marginal costs equal marginal returns. The returns to reintegration activities are not only reduced threats of getting a sanction, but if reintegration activities are effective, also a higher probability of earlier work resumption and hence reduced sick pay outlays. If optimal reintegration effort already exceeds the minimum requirements imposed by the NSII, introducing stricter screening does not change the employer's behavior. However, if optimal effort is below the minimum requirements, stricter screening induces employers to increase reintegration activities. Stricter screening then leads to more reintegration activities and if these activities are effective, work resumption rates during sickness absenteeism should increase and DI application rates decrease. This is the direct effect of stricter screening.

Stricter screening also reduces the attractiveness of the DI program to potential applicants and may trigger a mechanism of self-selection or *self-screening* (Parsons 1991). The decision to start a DI application process involves a comparison of expected utilities of alternatives, such as unemployment, early retirement and continuing work. Stricter screening increases the costs of applying for a DI benefit. Self-screening means that potential applicants who think that their DI application might not meet the eligibility requirements choose not to apply for the program. Indeed, Parsons finds that self-screening is important in explaining fluctuations in the inflow into DI in the United States.

If self-screening is important, we should observe fewer 13-week sickness reports in regions where a stricter screening policy is implemented. If direct effects of screening are important, we should observe more workers who report sick at week 13 to resume work before they file a DI claim (week 39). Recall that reintegration plans are drafted after eight weeks of sickness absence. Therefore, the impact of the reintegration activities on work resumption will mainly be concentrated after 13 weeks. A possible consequence of self-screening is that workers who without stricter screening would have applied for the DI program will apply instead for other programs, such as UI. Riphahn (1997) finds evidence for substitution between UI and DI schemes in Germany. If this is also relevant here, we should expect to find higher UI inflow rates in the treatment regions.

The effect of stricter screening on the sanction rate is ambiguous. In the treatment regions noncompliance with the minimum requirements is more likely to be detected. This implies more sanctions due to stricter screening. However, self-screening and increased reintegration activities due to the stricter screening are expected to reduce noncompliance in the treatment regions. It is unclear which of these effects dominates.

The start of the experiment in January 2003 coincides with the start of screening DI applications according to the new gatekeeper protocol. Therefore, at that time employers, employees, and occupational health agencies in both the treatment and the control regions did not have any experience with the way in which reintegration activities would be judged. Therefore, at the start of the experiment, employers in the treatment and in the control regions were expected to display the same behavior. However, sanction probabilities in the treatment regions were higher and therefore employers were notified faster about noncompliance. Employers in the treatment regions should thus learn sooner about the minimum reintegration activities requirements. The occupational health agencies serve many different employers and collect and transmit information about changes in scrutiny at the regional offices of the NSII. Therefore, employers not only learn through their own experience, but also through information disseminated by the occupational health agencies.

The experiment was not continued after 2003 and we do not have data later than 2003. As will be discussed in the next section, we observe per year how many workers of a specific type apply for DI benefits (or are reported 13-week sick at the NSII), but not the exact date. The larger part of the DI applications that are observed in our data are from workers who reported sick in 2002 or at the start of 2003 when employers and workers had not yet experienced the details of the new screening policy. This implies that self-screening at the start of the DI application process (if present) is most likely not yet reflected in the DI application rates of 2003. Reductions in the 2003 DI application rates due to stricter screening should therefore be interpreted as *direct effects* of the increased reintegration activities on work resumption. The 13-week absence reports are much less affected by the direct effects of increased reintegration activities and therefore the effect of the stricter screening on these reports should be interpreted as the consequence of *self-screening*. However, also the 2003 reports on 13-week sickness absenteeism include reports from workers who became sick at the end of 2002 or the very beginning of 2003 when information about the scrutiny of the screening process was not available. Therefore, the estimated self-screening effect as measured by the reduction in 13-week sickness reports will most likely be an underestimate of the long-run effect of stricter screening.

The distinction between direct effects and self-screening may not be as clear as mentioned above. First, reintegration activities may already be effective before the 13th week of sickness absence and differences in 13-week sickness reports may therefore also partly be due to direct effects. Recall that reintegration activities are only drafted in a report after eight weeks of sickness. We therefore expect this direct effect to be modest. Second, and perhaps more important is that also after week 13 sick workers may self-screen, for example because they realize that reintegration activities are more intense than expected. On the other hand, sickness absence induces considerable costs

to the employer. Employers may not only suffer productivity losses and pay for sickness benefits, they should also pay for the services of occupational health agencies. Workers too may suffer from prolonged sickness absences beyond week 13 because they may jeopardize their future prospects at the firm. Most self-screening is therefore likely to take place before or early in the period of sickness absence.

5. Data

Our data are from the administrative records of the NSII and describe the period 2001–2003. For each year, the data contain all insured individuals in 45 sectors of industry, which cover about 50% of the Dutch work force.[5] Not included are branches like construction, retail, health care, and civil service. Stratified by age, gender, sector of industry, and region, we observe for 2001, 2002 and 2003 (i) the number of insured individuals on January 1, (ii) the number of 13-week sickness reports during the year, (iii) the number of DI applications at week 39 during the year, and (iv) the number of individuals flowing into UI during the year. For instance, the first row of our data shows that in 2001 in the first region there were 104 males below 25 working in sector 5 of which four reported sick, no one applied for DI, and six entered UI in 2001. This setup of the data does not allow us to identify specific individuals, but only groups of individuals sharing the same observed characteristics. Therefore we do not have a true panel data structure at the individual level. Also we do not know if the same individuals who reported sick later entered DI or UI and we do not know the exact date at which an event happened.

In total our data contain about 3.3 million insured individuals for each year. On average each region thus contains about 125,000 insured workers. However, the smallest region has about 50,000 insured individuals, while the largest region includes almost 200,000 insured individuals. Table 2 shows that on average the treatment regions have the same size as the control regions. Except for one (control) region, in all regions the number of insured individuals decreased between 2001 and 2003 (by at most 4.4% in a control region). The two treatment regions experienced a reduction of 2.0% and 2.7%, respectively, in insured workers. The reduction in insured workers is the consequence of low GDP growth prevailing in our observation period. The 13-week sickness absence rate and DI applications rate decreased, while the inflow into UI increased over time. The treatment regions do not differ much from the control regions in terms of sickness absenteeism, DI applications and UI inflow rates. The sanction rate is only observed in 2003, since sanctions were only introduced when the gatekeeper protocol became effective. The data on sanctions describe the total number of sanctions given in each region. Therefore, we cannot link sanctions to specific

5. Traditionally, the NSII consisted of five different organizations each handling a number of sectors of industry. In 2002 these organizations were officially integrated, but in 2003 caseworkers were still connected to one of the original organizations and databases of the different organizations were not yet merged. In the experiment, we only instructed the caseworkers of the largest original organization handling 45 industries.

TABLE 2. Some descriptive statistics of the data.

	2001		2002		2003	
	Treatment	Control	Treatment	Control	Treatment	Control
Number of insured individuals	131,560	128,725	131,330	128,494	128,388	125,880
Sickness absenteeism rate (13 weeks)	0.0504	0.0477	0.0482	0.0448	0.0415	0.0403
DI application rate	0.0140	0.0153	0.0138	0.0139	0.0101	0.0107
UI inflow rate	0.0256	0.0318	0.0423	0.0444	0.0574	0.0601
Sanction rate					0.0408	0.0528
Fraction female	0.286	0.312	0.290	0.315	0.289	0.315
Age <35	0.177	0.165	0.176	0.164	0.170	0.158
Age 35–44	0.292	0.299	0.280	0.288	0.268	0.277
Age 45–54	0.262	0.268	0.267	0.272	0.273	0.277
Age ≥55	0.192	0.195	0.192	0.195	0.195	0.198

The sickness absenteeism rate, DI application rate and UI inflow rate are fractions of the number of insured individuals. The sanction rate is a fraction of the number of DI applications.

individuals. The sanction rate is the number of sanctions as a fraction of the number of DI applications. Sanction rates are on average higher in control regions than in treatment regions. The fraction of females remained constant over time, but over the years the insured population aged. The latter may be a combination of an ageing work force, reduced early retirement options (see OECD 2004, for trends on early retirement in the Netherlands) and fewer job opportunities for school leavers and young workers. Treatment and control regions are also very similar in terms of the share of females and age structure of the insured population.

As mentioned in Section 3, treatment regions were not randomly selected, but assigned by the NSII. The NSII tried to choose two treatment regions, which were representative for all regions in terms of DI inflow rates, size of insured population, and sectors of industry. It is however important to check if the choice of treatment regions is unrelated to relevant outcome measures. Recall that stricter screening of reintegration reports started in January 2003. Therefore, we can use the pre-experiment years 2001 and 2002 to investigate random assignment. In Table 3 we show for 2002 sample means and standard errors of the outcome variables in treatment and control regions.

TABLE 3. Pre-experiment outcomes in treatment and control regions (year 2002).

	Treatment Regions	Control Regions
Sickness absenteeism (13 weeks)	0.04823	0.04483
	(0.00079)	(0.00433)
DI applications	0.01376	0.01392
	(0.00011)	(0.00300)
UI inflow	0.04228	0.04436
	(0.00542)	(0.09167)

Standard errors in parentheses.

TABLE 4. Test for (conditional) random selection of treatment regions: difference-in-difference estimates (2002 compared to 2001).

	Sickness absenteeism	DI applications	UI inflow
Trend	−0.00291	−0.00142	0.01257
	(0.00018)	(0.00023)	(0.00106)
Treatment region	0.00072	0.00113	0.00041
	(0.00063)	(0.00082)	(0.00380)
Observations		26 regions	

Standard errors in parentheses.

Treatment regions do not differ substantially with respect to the outcome variables from control regions, although 13-week sickness is somewhat higher and UI inflow is slightly lower. The main goal of stricter screening is to reduce the DI application rate. Since the table shows that DI application rates were virtually identical, we conclude that the treatment regions were not selected on their DI application rates (or any other relevant outcome) in 2002.

Since our data describe multiple years, we only need random selection of treatment regions conditional on regional fixed effects. A simple test for conditional random assignment is to check whether in treatment regions outcome variables changed differently between the pre-experiment years 2001 and 2002 than in control regions. More specifically, we take the panel data model at the level of the region for the years 2001 and 2002:

$$Y_{r,t} = \alpha + \beta Year_t + \gamma Treatment_{r,t} + \delta_r + u_{r,t}, \tag{1}$$

where β describes the nationwide time trend, $Treatment_{r,t}$ is an indicator that only equals 1 in the treatment regions in the year 2002 and δ_r is the region-specific effect. We estimated this equation in first differences:

$$Y_{r,2002} - Y_{r,2001} = \beta + \gamma Treatment_r + e_r. \tag{2}$$

The estimation results in Table 4 show that the outcomes in the treatment regions do not significantly differ from those in the control regions. We therefore conclude that the selection of treatment regions is exogenous with respect to pre-experiment outcomes. But as we do not have any information prior to 2001, we cannot rule out that regions differed prior to 2001.

6. Results

In this section we first discuss the effects of stricter screening on 13-week sickness reports and DI applications. Next, we investigate how the sanction rate is affected by stricter screening. We then report on some additional analyses and discuss the issue of targeting efficiency. We finally report on results of a simple cost–benefit analysis.

6.1. Effects on Sickness Absence and DI Applications

Our main parameters of interest are the effects of stricter screening on 13-week sickness absenteeism reports and DI applications. Recall from Section 4 that our data only allow us to measure this effect during the first year after stricter screening was introduced. Therefore, we can only focus on short-run effects of stricter screening.

In Table 5 we present the results from difference-in-difference analyses. For each region we use as outcomes the number of 13-week sickness absenteeism reports and the number of DI applications as a fraction of the insured population in the pre-experiment year 2002 and the experiment year 2003. Effectively we estimate a regression equation similar to equation (2), but for the years 2002 and 2003. The trend effects show that both the probability that a worker reports sick and the probability that a worker applies for DI benefits decrease significantly between 2002 and 2003. This may be the result of the nationwide regime shift that occurred in April 2002 and became effective in January 2003. But it might as well be the result of adverse macroeconomic conditions. Absence rates are found to be pro-cyclical (e.g. Johansson and Palme 1996) and the Dutch economy was in a recession during our observation period. From the table we see that stricter screening causes a decrease in sickness absenteeism reports and in DI applications. However, only the effect on 13-week sickness reports is significant.

The key assumption of the difference-in-difference analysis is that the time trend is the same in all regions. Because our data also contain information on the year 2001, we can relax this assumption. In particular, we can allow for region-specific time trends by extending our model specification to

$$Y_{r,t} = \alpha + \beta_r Year_t + \gamma Treatment_{r,t} + \delta_r + u_{r,t}. \tag{3}$$

The parameters β_r describe regional time trends. We can eliminate both the regional time trends and the region-specific effects δ_r by taking the differences twice:

$$(Y_{r,2003} - Y_{r,2002}) - (Y_{r,2002} - Y_{r,2001}) = \gamma Treatment_r + e_r. \tag{4}$$

We estimate this model using OLS (omitting the intercept). We find that the effect of stricter screening on 13-week sickness reports is −0.00449 with a standard error of 0.00156. The estimated treatment effect on DI applications is −0.00339 with a

TABLE 5. The effect of stricter screening: regional difference-in-difference estimates (2003 compared to 2002).

	Sickness absenteeism	DI applications
Trend	−0.00455	−0.00320
	(0.00028)	(0.00021)
Treatment (stricter screening)	−0.00213	−0.00048
	(0.00103)	(0.00076)
Observations	26 regions	

Standard errors in parentheses.

standard error of 0.00145. Both estimated effects are thus significant and larger than in a model that does not allow for regional differences in time trends.

The analyses at the regional level above are similar to between-group estimates. The estimator for γ is most efficient if the errors e_r are homoskedastic, which is according to Donald and Lang (2007) a natural assumption given the very large number of insured workers in each region.[6] The regional level analyses do not exploit individual characteristics which are available in the data. Taking account of individual heterogeneity is attractive for two reasons. First, it explains some of the variance in the error terms at the regional level, which improves the precision of estimates. Second, we can investigate whether effects of stricter screening differ between individuals with different characteristics. We specify Logit models for the individual probability of 13-week sickness and the probability to apply for DI benefits after 39 weeks:

$$\begin{aligned}\Pr(Y_{i,r,t} = 1) = \Lambda(\delta_r + \beta_1 Age_{i,t} \times Gender_i + \beta_2 Year_t + \beta_3 Sector_{i,t} \\ + \beta_4 Year_t \times Sector_{i,t} + \gamma Treatment_{r,t}).\end{aligned} \tag{5}$$

$\Lambda(.)$ stands for the Logit function. Other than the indicator for stricter screening, we include as regressors regional indicators, gender of the individual interacted with age dummies, a nationwide time trend, sector dummies, and sector dummies interacted with time trends. The latter set of variables allows for different sector trends between 2002 and 2003. Recall that our data only reveal that there are $N_{i,r,t}$ individuals who live in region r in year t and have characteristics $Age_{i,t}$, $Gender_i$, and $Sector_{i,t}$. Out of these $N_{i,r,t}$ individuals $n_{i,r,t}$ are observed to make a 13-week sickness report (or similarly apply for DI benefits), which yields the loglikelihood function

$$Log\ell = \sum_{i,r,t} n_{i,r,t} \times \log(\Lambda(\cdot)) + (N_{i,r,t} - n_{i,r,t}) \times \log(1 - \Lambda(\cdot)). \tag{6}$$

To estimate this model we use only data for 2002 and 2003. The estimated effect of stricter screening on 13-week sickness reports is −0.05622 with the clustered standard error 0.02689 and the estimated effect on DI applications is −0.04899 with a clustered standard error 0.01970.

Donald and Lang (2007) show that clustered standard errors can be unreliable if the number of groups is small. Their Monte Carlo simulations show that clustered standard errors are much too small in the 2×2 case. Donald and Lang propose a two-step estimator with better small-sample properties. Our data contain 26 regions, which is small but much more than the two for which Donald and Lang show simulation results. It is therefore interesting to also consider the two-step estimator of Donald and Lang.[7] With this estimator we find that the effect of stricter screening on sickness absenteeism is −0.06339 (with a standard error of 0.00716) and on DI applications

6. Weighting regional observations by the number of insured individuals in the region does not affect the results.

7. We implemented the two-step estimator by first estimating the Logit model with fixed effects for each region in each year. Next, we decompose the estimated region–year fixed effects into a time trend, regional fixed effects, and a treatment effect.

TABLE 6. The effect of stricter screening: marginal effects from individual Logit estimates.

	Sickness absenteeism	DI applications
Trend	−0.00421	−0.00074
	(0.00212)	(0.00088)
Treatment (stricter screening)	−0.00223	−0.00058
	(0.00107)	(0.00023)
Male <35	–	–
Female <35	0.02286	0.01028
	(0.00047)	(0.00026)
Male 35–44	0.02383	0.00792
	(0.00032)	(0.00023)
Female 35–44	0.04072	0.01594
	(0.00041)	(0.00023)
Male 45–54	0.03467	0.01286
	(0.00031)	(0.00028)
Female 45–54 years	0.04680	0.01926
	(0.00049)	(0.00031)
Male ≥55	0.04403	0.01436
	(0.00046)	(0.00026)
Female ≥55	0.04835	0.01861
	(0.00061)	(0.00036)
Observations	6,624,402 individuals	

In this regression is also controlled for region, sector and interactions between the sector and the time trend (in total 114 additional parameters). Standard errors (corrected for clustering at the regional level) in parentheses.

−0.04756 (standard error 0.01746). The estimated effects are thus similar to the effects directly estimated in the Logit specification.

Estimated coefficients from Logit models are not easy to interpret. Therefore, we focus on marginal effects of stricter screening:

$$\frac{1}{\sum_{i,r,t} N_{i,r,t}} \sum_{i,r,t} N_{i,r,t} \times \hat{\Lambda}(\cdot) \times (1 - \hat{\Lambda}(\cdot)) \times \hat{\gamma}. \tag{7}$$

The marginal effects we report are thus the average over all individual marginal effects. Standard errors of the marginal effects are computed using the Delta method and corrected for clustering within regions. Table 6 shows the marginal effects from the estimated Logit models. We see little changes in the effect of stricter screening when we compare the difference-in-difference estimates with the estimates of the Logit model. This confirms that selection of treatment regions was not correlated to changes in the composition of the insured population over time. The standard errors for DI applications in the Logit model are smaller than those in the difference-in-difference model. The effect of stricter screening is significant at the 5% level, both for 13-week sickness reports and DI applications. We also find significant age and gender effects. According to Table 6, the reference group has the lowest absence and DI application rates. Sickness absence and DI application are higher for females and increase with age. We return to these age and gender effects in Section 6.3 when we present some sensitivity analyses.

TABLE 7. Simulations for sickness absence and DI applications under different screening scenarios applying to 2003 using the estimates from the Logit model.

	Sickness absenteeism	DI applications
Predicted in current situation	130,425	34,065
Without nationwide stricter screening	131,006	34,192
With nationwide stricter screening	124,164	32,564

Recall from Section 4 that we can distinguish two effects of stricter screening. As a direct effect, stricter screening may increase reintegration efforts and work resumption rates among sick-listed workers. And, indirectly, reduced prospects may induce workers and employers not to start a DI application process. We argued in Section 4 that the effect of stricter screening on DI applications reflects direct effects, while the effect on 13-week sickness reports reflects self-screening effects. The empirical results show that stricter screening reduces both the incidence of 13-week sickness absenteeism and the number of DI benefit applications, which implies that both effects are important. To provide some more information on the quantitative importance of both effects, we provide in Table 7 the predicted number of sickness absenteeism reports and DI applications in 2003. Furthermore, Table 7 includes what the number of 13-week sickness reports and DI applications would have been if stricter screening would have been implemented either nationwide, or nowhere in the Netherlands. From the table we can see that nationwide implementation of stricter screening would reduce the number of sickness absenteeism reports by 5.2% and the number of DI applications by 4.8%. To the extent that our interpretation of both effects is valid, this implies that indirect self-screening effects and direct effects of increased reintegration efforts are quantitatively equally important.

6.2. Effects on Sanctions

The gatekeeper protocol provides the NSII with the possibility of sanctioning employers. Since these sanctions did not exist prior to 2003, we cannot control for region-specific unobserved effects. This reduces the reliability of the estimated effects of stricter screening on sanctions. The data on sanctions are also less informative than data on other outcome variables; we only observe the total number of sanctions applied in each region and we cannot connect these sanctions to specific DI applications. Therefore, we should interpret the empirical results with care and we should not draw too strong conclusions about how stricter screening affects the sanction rate.

Table 8 provides the results from an OLS regression on the sanction rate in each region. We control for regional differences by including lagged (2002) 13-week sickness report rates, lagged DI benefit application rates, and lagged UI inflow rates. The estimation results show that the sanction rate is slightly lower in the treatment regions but this effect is not significant. However, sanction rates are lower in regions with a high DI application rate and a high UI inflow rate. A propensity score weighting

TABLE 8. The effect of stricter screening on sanctions: results from an OLS regression on the sanction probability in 2003.

	Sanctions
Intercept	0.0648 (0.0483)
Treatment (stricter screening)	−0.0064 (0.0175)
Sickness absenteeism reports in 2002	−0.7398 (1.4198)
DI applications in 2002	−4.1712 (1.8475)
UI inflow in 2002	1.7875 (0.6013)
R^2	0.383
Observations	26 regions

Standard errors in parentheses.

estimator shows the same result. Using the estimator presented in Hirano, Imbens, and Ridder (2003) we find an effect of stricter screening on the sanction rate of −0.0298 with a standard error of 0.4864.[8] The insignificant effect does not contradict our previous finding. Recall that stricter screening causes two mechanisms affecting the sanction rate in opposite directions. First, stricter screening has the direct effect that more cases of noncompliance will be detected. And second, if stricter screening is anticipated, then there will be an indirect effect in that the composition of the DI applicants shifts towards applicants who undertake more reintegration activities and are thus less likely to be sanctioned.

6.3. Sensitivity Analyses

The Logit estimates of Section 6.1 show that sickness risks and DI application probabilities vary with individual characteristics. Recall from Section 2.1 that incentives to apply for DI vary with age. Therefore, also the effect of stricter screening may vary with age of the applicant. We extend the Logit model in equation (5) such that the effect of stricter screening γ can depend on age and gender. Again we compute marginal effects of stricter screening for individuals with different observed characteristics. The marginal effects reported in Table 9 indeed show that there are differential age and gender effects. Stricter screening affects the 13-week sickness report rates of males older that 35. However, for males above age 55 the effect is only significant at the 10% level. For females we find that until age 55 the effects of stricter screening on 13-week sickness reports are larger than for males, but the effect is only significant for the group aged 35–44. For the DI applications, stricter screening only affects female DI application rates and the effects increase with age.

To understand these results one should consider the composition of the DI inflow. Compared with other OECD countries, the inflow into the Dutch DI program is characterized by relatively high fractions of women and young workers (OECD 2003).

8. When applying this method the propensity score was specified as a Logit model and included the same regressors as the OLS regression of Table 8.

TABLE 9. Effects of stricter screening by age and gender: marginal effects from individual Logit estimates.

	Sickness absenteeism	DI applications
Stricter screening interacted with:		
Male <35	−0.00077	0.00103
	(0.00162)	(0.00055)
Female <35	−0.00254	−0.00127
	(0.00141)	(0.00079)
Male 35–44	−0.00282	0.00009
	(0.00075)	(0.00031)
Female 35–44	−0.00302	−0.00147
Female 35–44	(0.00191)	(0.00080)
Male 45–54	−0.00186	−0.00023
	(0.00093)	(0.00057)
Female 45–54	−0.00362	−0.00173
	(0.00257)	(0.00073)
Male ≥55	−0.00241	−0.00059
	(0.00131)	(0.00099)
Female ≥55	−0.00030	−0.00458
	(0.00429)	(0.00101)
Observations	6,624,402 individuals	

In this regression is also controlled for a time trend, gender interacted with age, region, sector and interactions between the sector and the time trend (in total 122 additional parameters). Standard errors (corrected for clustering at the regional level) in parentheses.

Until the mid-1980s, women had lower DI inflow rates than men, but since 1985 female inflow rates exceed those for men. At the end of the 1990s, female inflow rates were almost twice the male inflow rates. Between 1991 and 2001 the number of male DI recipients decreased by 13%, while the number of female beneficiaries increased by 43%. The sharp increase in female disability benefit recipiency was accompanied by a strong growth of labor force participation of mothers, which was traditionally low. Whereas three out of four women stopped working after the birth of their first child in the 1970s, only one third stopped 20 years later. Part of this shift was accommodated by DI benefits which allow market production to be replaced by home production without a sharp drop in household income (De Jong 2004).[9] Bratberg, Dahl, and Risa (2002) examine whether career and family obligations increase sickness absence and DI inflow rates among mothers and conclude that it does. Among the OECD countries, only in the Scandinavian countries is the fraction of female workers in the inflow in DI as high as in the Netherlands.[10] However, Dutch female labor force participation rates are much lower than in the Scandinavian countries.

9. One could see if the pattern differs between married and unmarried women and between women with and without children. In our data we do not observe the marital status, nor do we have information on the presence of children.

10. Like in the Netherlands, the Scandinavian DI schemes are also characterized by relatively lenient eligibility conditions and high replacement rates.

TABLE 10. The effect of stricter screening for both treatment regions: marginal effects from individual Logit estimates.

	Sickness absenteeism	DI applications
Apeldoorn	0.00001	−0.00056
	(0.00022)	(0.00023)
Hengelo	−0.00414	−0.00059
	(0.00043)	(0.00029)
Observations	6,624,402 individuals	

In this regression is also controlled for a time trend, gender interacted with age, region, sector and interactions between the sector and the time trend (in total 122 additional parameters). Standard errors (corrected for clustering at the regional level) in parentheses.

Recall from Table 1 that the caseworkers in both experiment regions used different methods to increase screening intensity. In the Apeldoorn region caseworkers arranged more face-to-face contacts with workers, while in the Hengelo region the caseworkers visited employers more often. The effect of stricter screening may vary depending on whether an employer or worker is checked. We therefore estimated the previous Logit model but now allowing for different effects for Apeldoorn and Hengelo. These coefficients, with standard errors adjusted for clustering at the regional level, are reported in Table 10. We find strong and significant effects on 13-week sickness reports for Hengelo, but not for Apeldoorn. The effect is equivalent to a reduction of about 10% in 13-week sickness reports in Hengelo. This might suggest that employer visits are the most effective screening method. However, the effects on DI applications do not differ between both regions and are in size similar to the results in Table 6.

6.4. Targeting Efficiency

Stricter screening leads to a fall in DI applications due to direct effects and indirect (self-screening) effects. Parsons (1991) argues that increases in the screening policy parameters, such as the waiting time before applying for DI and the denial rate, can lead to greater reductions in DI applications among disabled than among able workers. This reduced targeting efficiency may be due to differences in credit market constraints, mortality expectations, and risk aversion between disabled and able workers. Parsons (1991) concludes for the United States that the change in the denial rate due to the 1977 disability funding crises did not lead to perverse self-screening effects.

Testing targeting efficiency requires health information on workers in different regions in different years. Unfortunately, our data do not contain such information. Therefore, we have to rely on other information to assess targeting efficiency. In the Dutch context it is not directly clear that stricter screening may lead to perverse effects. Unlike the United States where individuals have no income during the DI waiting period, Dutch workers get sick pay up to 100% of their net earnings during the entire one year waiting period. Stricter screening of reintegration reports as such can lead to an extension of the waiting period if a sanction is applied, but it is primarily the employer who is held responsible (see Section 2). So, stricter screening does not have

direct negative financial consequences for workers, unless employers penalize sick employees who are unwilling to collaborate with a reasonable reintegration plan. This should affect the less disabled more than those with severe disabilities. For workers there may also be indirect financial consequences of reporting sick and trying to get into a DI program. For instance, workers with a history of long-term sickness absenteeism may have lower promotion rates and higher firing rates. These costs may in fact be higher for able workers than for truly disabled workers.

It is mainly the employer who suffers financially from stricter screening. To avoid a punitive extension of the waiting period, the employer must convincingly argue why reintegration activities did not lead to work resumption during the waiting period. It is much more difficult to argue that reintegration activities are ineffective when the worker is not truly disabled. So the employer faces a higher sanction risk when using DI to lay off able workers.

Older workers' labor force participation rates are much lower in the Netherlands than for example in the United States. Burkhauser et al. (1999) compare health, work and economic well-being of older men and women in the United States and the Netherlands and conclude that differences between Dutch and U.S. institutions, rather than differences in health explain the differences in labor force participation rates. If we combine these institutional facts with the results from the sensitivity analyses discussed in the previous section, where we saw that the impact of stricter screening is stronger for workers aged 45 years and older, we conclude that stricter screening most likely reduces moral hazard among older workers.

If stricter screening would scare truly disabled workers away from applying for DI benefits, we would expect these workers to leave the firm through some other channel. Able workers who do not apply for DI benefits can continue working in the firm, while this is much more difficult for disabled workers. Therefore, if stricter screening would have perverse effects in the sense that truly disabled workers would not apply for DI, we should expect to see an increase in the UI inflow rate. It is not uncommon that there are interactions between UI and sickness and DI programs (e.g. Riphahn 1997; Larsson 2006).

Our data are informative on UI inflow. Table 11 presents estimates for the marginal effects of the Logit model for the effect of stricter screening on the UI inflow. We do not find an increase in UI inflow in the treatment regions; the estimates of the stricter screening are even negative. This result remains if we allow for different effects for different age and gender groups (see Table 12). The lack of spillovers between DI and UI suggests that the workers who did not apply for DI benefits did not flow into UI and most likely remained at work. This again should be taken as evidence that stricter screening did not harm the disabled workers, but rather reduced the attractiveness of the DI program for those able to work.[11]

11. Individuals insured for DI are almost always also insured for UI (both are compulsory insurances in the Netherlands). It should be noted, however, that the youngest workers may not qualify for UI when they have worked less than 39 weeks. This is a very small group of the total group of young workers. Furthermore, a share of the oldest workers (55+) may have access to alternative labor force exit routes such as early retirement.

TABLE 11. The effect of stricter screening: marginal effects from individual Logit estimates.

	UI inflow
Trend	0.02354
	(0.00327)
Treatment (stricter screening)	−0.00043
	(0.00121)
Male <35	–
Female <35	0.01547
	(0.00091)
Male 35–44	0.00366
	(0.00095)
Female 35–44	0.02418
	(0.00088)
Male 45–54	0.00470
	(0.00144)
Female 45–54	0.02240
	(0.00152)
Male ≥55	0.01614
	(0.00187)
Female ≥55	0.02801
	(0.00220)
Observations	6,624,402 individuals

In this regression is also controlled for region, sector and interactions between the sector and the time trend (in total 114 additional parameters). Standard errors (corrected for clustering at the regional level) in parentheses.

TABLE 12. The effects of stricter screening by age and gender on UI inflow: marginal effects from Logit estimates.

	UI inflow
Stricter screening interacted with:	
Male <35 years	0.00001
	(0.00128)
Female <35 years	−0.00015
	(0.00196)
Male 35–44 years	−0.00063
	(0.00148)
Female 35–44 years	−0.00066
	(0.00142)
Male 45–54 years	−0.00053
	(0.00309)
Female 45–54 years	−0.00044
	(0.00259)
Male ≥55 years	−0.00141
	(0.00223)
Female ≥55 years	−0.00118
	(0.00247)
Observations	6,624,402 individuals

In this regression is also controlled for a time trend, gender interacted with age, region, sector and interactions between the sector and the time trend (in total 122 additional parameters). Standard errors (corrected for clustering at the regional level) in parentheses.

6.5. Back-of-the-envelope Calculation of Monetary Costs–Benefits

We can use our estimation results to perform some back-of-the-envelope calculations of stricter screening. It should be stressed that we only look at the direct monetary costs and benefits and ignore wider social costs and benefits. We do this in order to provide some more insight into the quantitative importance of screening on DI benefit payments. We use the results of Table 7 to compare a situation where stricter screening is applied nationwide to a situation where the standard screening procedure is applied across the country. The results in Table 7 apply to the year 2003. Without stricter screening each caseworker handles on average 110 DI applications per year. For all 34,196 applications about 311 caseworkers were needed. Stricter screening requires an additional time investment per caseworker of 9.4% of the total time spent on a DI application. But given the reduction in the annual number of applications due to stricter screening only 13 additional caseworkers are required, which costs approximately €1,320,000 per year.

Stricter screening induces a reduction of 1,633 DI applications per year. Recall that an application is made after 39 weeks of sickness absenteeism, while DI benefit entitlement starts after 52 weeks of sickness absenteeism. Due to both benefit denials and recoveries between 39 and 52 weeks, only 47% of the applications lead to actual enrollment into the DI program. The average DI benefit is €11,745 per year and average benefit duration is 12.9 years.[12] If we use a discount rate of 5% per year, the expected present value of future DI benefit payments is slightly less than €85,000. Hence, the total amount of averted DI benefit payments by stricter screening are around €64,800,000.

This monetary cost–benefit analysis assumes that a nationwide introduction of stricter screening would not affect the composition of the inflow into DI. In particular, annual exit probabilities are assumed to remain unchanged and the average benefit level is assumed not to change. It is, however, likely that stricter screening shifts the population flowing into DI towards individuals who have worse health and labor market prospects and thus on average longer DI spells. This reduces the expected monetary benefits of stricter screening. On the other hand, the cost–benefit analysis does not take into account that stricter screening also significantly reduces the 13-week sickness report rate, which reduces administrative costs. The main conclusion of these rough calculations is that the monetary costs of implementing the procedure of stricter screening are only a small fraction of the monetary benefits.

7. Conclusions

This paper looks at effects of stricter screening of applications for DI in the Netherlands. In particular, we examine whether stricter screening affects the prevalence of

12. The gross annual exit probabilities from DI are in the first year 0.16, in the second year 0.17, in the third year 0.10, in the fourth year 0.07, and around 0.06 afterwards.

long-term sickness absence and DI benefit applications. With little screening, workers and employers can decide to devote none or only minimal effort to try and get the worker back to work. Stricter screening is a policy measure to reduce such moral hazard as it forces employers and workers to increase reintegration effort. This reduces the attractiveness of the DI program and therefore stricter screening may also trigger a mechanism of self-selection by potential applicants. Our empirical strategy was to exploit exogenous regional variation in the intensity of the screening of DI applications. In particular, we designed an experiment in which we instructed caseworkers at two local DI offices to implement a regime of stricter screening of applications than in the rest of the country.

We use data from the administrative records of the NSII and find that stricter screening causes both self-selection and increased work resumption rates during sickness absenteeism and that both effects are quantitatively equally important. We argue that stricter screening mainly reduces the DI application rate of able workers and thereby reduces moral hazard and improves targeting efficiency. A crude cost–benefit analysis shows that the monetary costs of stricter screening are only a small fraction of the monetary benefits.

This study shows that variations in the screening procedure can have an important impact on the inflow into targeted transfer programs. Since the monetary costs of screening are relatively low, stricter screening may be an attractive policy instrument to reduce moral hazard. An important condition for stricter screening is that the social program should have well-defined eligibility criteria. Furthermore, stricter screening should be accompanied by a sanction scheme which contains substantial financial penalties for noncompliance. Programs covering wage loss due to unemployment also screen applications and sanction for unnecessary job loss. Our findings may therefore not only have implications for DI programs around the world but also for other targeted transfer programs.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Stricter Screening data file (STATA .dta file)
Appendix S2. Stricter Screening data file readme (txt document)
Appendix S3. Work (.do document)
Appendix S4. Output (.log file)

Please note: Blackwell Publishing are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

Autor, David H. and Mark G. Duggan (2003). "The Rise in the Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics*, 118, 157–205.

Bound, John and Richard V. Burkhauser (1999). "Economic Analysis of Transfer Programs Targeted on People with Disabilities." In *Handbook of Labor Economics, Volume 3C*, edited by O. C. Ashenfelter and D. Card. North-Holland.

Bratberg, Espen, Svenn-Åge Dahl, and Alf Erling Risa (2002). "The Double Burden—Do Combinations of Career and Family Obligations Increase Sickness Absence among Women?" *European Sociological Review*, 18, 233–249.

Burkhauser, Richard V., Mary C. Daly, and Philip de Jong (2008). "Curing the Dutch Disease: Lessons for United States Disability Policy." Working paper. Cornell University.

Burkhauser, Richard V., Debra Dwyer, Maarten Lindeboom, Jules Theeuwes, and Isolde Woittiez (1999). "Health, Work and Economic Well-being of Older Workers, aged 51 to 61: a Cross-national Comparison using the United States HRS and the Netherlands CERRA Data Sets." In *Wealth, Work, and Health: Innovations in Measurement in the Social Sciences*, edited by James Smith and Robert Willis. University of Michigan Press.

Chen, Susan and Wilbert van der Klaauw (2008). "The Effect of Disability Insurance on Labor Supply of Older Individuals in the 1990s." *Journal of Econometrics*, 142, 757–784.

De Jong, Philip (2004). "De WAO, 1967–2006—Over Opkomst en Ondergang van een Sociale Arbeidsongeschiktheidsverzekering." *Tijdschrift Politieke Ekonomie*, 26, 4–26. (In Dutch)

Donald, Stephen G. and Kevin Lang (2007). "Inference with Difference-in-differences and Other Panel Data." *Review of Economics and Statistics*, 89, 221–233.

Gruber, Jonathan (2000). "Disability Insurance Benefits and Labor Supply." *Journal of Political Economy*, 108, 1162–1183.

Gruber, Jonathan and Jeffrey D. Kubik (1997). "Disability Insurance Rejection Rates and the Labor Supply of Older Workers." *Journal of Public Economics*, 64, 1–23.

Halpern, Janice and Jerry A. Hausman (1986). "Choice under Uncertainty: a Model of Applications for the Social Security Disability Insurance Program." *Journal of Public Economics*, 31, 131–161.

Hirano, Kei, Guido W. Imbens, and Geert Ridder (2003). "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score." *Econometrica*, 71, 1161–1189.

Johansson, Per and Mårten Palme (1996). "Do Economic Incentives Affect Work Absence? Empirical Evidence Using Swedish Micro Data." *Journal of Public Economics*, 59, 195–218.

Johansson, Per and Mårten Palme (2002). "Assessing the Effect of a Public Policy on Worker Absenteeism." *Journal of Human Resources*, 37, 381–409.

Larsson, Laura (2006). "Sick of Being Unemployed? Interactions between Unemployment and Sickness Insurance." *Scandinavian Journal of Economics*, 108, 97–113.

OECD (2003). *Transforming Disability into Ability: Policies to Promote Work and Income Security for Disabled People*. Directorate for Employment, Labor and Social affairs, Section on Sickness and Disability Policies.

OECD (2004). *Economic Survey of the Netherlands: Reform of the Sickness and Disability Benefit Scheme*. Directorate for Employment, Labor and Social affairs, Section on Sickness and Disability Policies.

Parsons, Donald O. (1991). "Self-screening in Targeted Public Transfer Programs." *Journal of Political Economy*, 99, 859–876.

Riphahn, Regina (1997). "Disability, Retirement and Unemployment—Substitute Pathways for Labor Force Exit?" *Applied Economics* 29, 551–561.