Johansson, Per; Laun, Lisa; Laun, Tobias

Working Paper

Screening stringency in the disability insurance program

Working Paper, No. 2013:10

Provided in Cooperation with:
IFAU - Institute for Evaluation of Labour Market and Education Policy, Uppsala

Suggested Citation: Johansson, Per; Laun, Lisa; Laun, Tobias (2013) : Screening stringency in the disability insurance program, Working Paper, No. 2013:10, Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala

This Version is available at:
http://hdl.handle.net/10419/82259

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU’s objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P.O. Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

ISSN 1651-1166
Screening stringency in the disability insurance program

by

Per Johansson\textsuperscript{b}, Lisa Laun\textsuperscript{c} and Tobias Laun\textsuperscript{d}

2\textsuperscript{nd} May, 2013

Abstract
We propose a strategy for assessing how the inflow to the disability insurance program has been governed over time. We analyze the ex-ante health of individuals entering the program, compared to individuals not entering the program in the same year, by using ex-post mortality. Applying the strategy to Sweden, we find large variation in the relative health of new beneficiaries compared to non-beneficiaries over time. Some of the fluctuations correspond well to formal changes to screening stringency. However, we also find large variation in health during periods when no changes to formal eligibility criteria have been pursued.

Keywords: Disability insurance, Screening stringency, Proportional hazard model
JEL-codes: C41, I18, J14
# Table of contents

1 Introduction ........................................................................................................... 3

2 The model ............................................................................................................... 7

3 Empirical strategy .................................................................................................. 13

3.1 Estimation ......................................................................................................... 13

3.2 Data .................................................................................................................. 16

4 An application to Sweden ...................................................................................... 17

4.1 The Swedish disability insurance system .......................................................... 17

4.2 Results ............................................................................................................... 18

4.3 Alternative explanations ..................................................................................... 24

5 Conclusion ............................................................................................................. 26

References ............................................................................................................... 28
1 Introduction

The disability insurance program has become one of the largest income maintenance programs in modern welfare states.\textsuperscript{1} Autor (2011) shows that, in the U.S., the share of 25–64 year-olds receiving benefits from the Social Security Disability Insurance (SSDI) increased from 2.3 to 4.6 percent between 1989 and 2009, and the SSDI Trust Fund is projected to be exhausted between 2015 and 2018. Sweden has also experienced large growth in disability benefits recipiency. As shown in Figure 1, the share of 30–64 year-olds receiving disability benefits increased from 8 to 12 percent between 1985 and 2008, but the fluctuations in the entry rate were large. The reasons for the large fluctuations over time are not fully understood. Some of the changes during the 1990s coincided with changes to formal eligibility rules, but the large increase during the early 2000s and the subsequent decline since the mid 2000s cannot be attributed to any formal program changes. According to OECD (2009), the decline can be attributed to a gradual improvement in the way existing regulations were being implemented, following the restructuring of the Social Insurance Agency from 21 regional offices to one central authority in 2005.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{The share of individuals receiving disability benefits and the disability benefit entry rate in ages 30–64 in Sweden, 1985–2008}
\end{figure}

\textsuperscript{1}See, e.g., Wise (2012) for a detailed description of disability insurance programs in several developed countries.
Policy makers have two main instruments for governing the inflow to the disability insurance program. Program generosity affects the attractiveness of the program and can be modified through changes in the compensation level or the cap on benefits. Program availability affects the likelihood of being awarded benefits and depends on screening stringency, which includes formal eligibility criteria as well as the implementation of program rules. While eligibility criteria are easily modified at the political level, the control of the implementation process is likely to be imperfect. Caseworker discretion and internal processes at the Social Insurance Agency are important in the eligibility determination process but are difficult for the insurance provider to monitor. Because of variation in implementation over time, the control of program inflow may change also during periods when no reforms at the political level are being pursued.

Empirical studies focusing on labor market effects of compensation levels and/or screening stringency have generally come to the conclusion that both policy instruments are important. However, the magnitude of the estimated effects varies to a large extent. A potential reason for the observed differences across time and countries is that the effects of changes in stringency can depend on the compensation level and, similarly, can the impact of changes in the compensation level depend on the current level of screening stringency. It is, therefore, important to be cognizant of both formal rules as well as their implementation to get a full understanding of why the magnitude of labor market effects induced by disability insurance reforms differ.

A key question in this context is whether fluctuations in the inflow to the disability insurance program can be explained by changes in the underlying health of the population, or whether they are due to changes in the control of program inflow. In this paper we propose a strategy for assessing how the inflow to the disability insurance program has been governed over time, by comparing the health of individuals admitted to the program to the health of individuals not admitted. Health, however, is difficult to measure. Self-rated health measures are subjective and likely to suffer from measurement error. Furthermore, self-rated health tends to be endogenous to labor supply choices, in the sense that indi-

See Kruger and Meyer (2002) for a review of several studies.
viduals receiving disability benefits are more prone to report an impaired work capacity.\(^3\) An objective measure of health that is accurately measured through population censuses and is difficult to fake is mortality. Information about mortality is also available in many countries and for long time periods, which facilitates comparisons across countries and over time. The immediate mortality of new disability beneficiaries is certainly low and not very informative. However, we analyze the ex-ante health of individuals entering the program, compared to individuals not entering the program in the same year, by using the ex-post mortality from all years in which the individuals are observed in the future. This approach allows us to control for changes in the underlying health when estimating fluctuations in program entry.

To structure the analysis, we also provide a stylized theoretical model of the application decision to the disability insurance program. We focus on screening stringency as a key determinant of program inflow. A potential indicator of screening stringency is the denial rate. Since Parsons (1991), however, it is well known that initial eligibility determination is an important self-screening mechanism. His results suggest that a 10 percent increase in the initial denial rate induced a 4 percent decrease in application rates after 2 years. Halpern and Hausman (1986) show that the probability of acceptance has a significant effect on the probability of applying for disability benefits. The denial rate is therefore an imperfect measure for studying the development of screening stringency over time. Also in our model we show that the composition of the applicant pool changes with screening stringency, and the effect of stringency on the denial rate is therefore undetermined. Going further, we show that the relative health of benefit applicants to non-applicants improves with a reduction in stringency.

The empirical analysis is performed using duration analysis, in which the differential development of health across cohorts is accounted for by estimating cohort-specific baseline mortality hazards. We apply the proposed strategy to the disability insurance program in Sweden, and study the development in general as well as across population groups. We find that changes in the mortality ratio of new disability beneficiaries compared to non-beneficiaries correspond well to the fluctuations in the disability benefit entry

\(^3\)See, e.g., Burkhauser, Daly, Houtenville and Nargis (2002).
rate over time. This suggests that the variation in program inflow cannot be accounted for by changes in the underlying health of the population. Changes to formal eligibility criteria are indeed reflected in the mortality ratio. Perhaps more interestingly, we also find fluctuations in the mortality ratio of new beneficiaries compared to non-beneficiaries during periods of no formal program changes. This suggests that also informal changes of the health assessment have impacted the inflow to the disability insurance program.

The association between population health and disability insurance recipiency has previously been assessed in Wise (2012). This paper provides a more formal framework for analyzing how program entry has been governed, after accounting for health changes. Our strategy captures the development of screening stringency due to changes in formal eligibility criteria as well as in implementation. The latter aspect has been difficult to assess in previous studies. Also other factors may affect the relative health of new disability beneficiaries compared to non-beneficiaries. As discussed above, program generosity is a different policy instrument for governing program inflow, but there are no changes in the benefit level during the studied period. The attractiveness of outside options may also be important for the application decision. The impact of labor market factors will be included in the estimates, and our measure can be taken as to capture the ability of the disability insurance program to withstand the influence of outside forces. A simple analysis suggests that the correlation between our estimate and a set of factors related to the attractiveness of outside options is weak. This suggests that our measure primarily captures changes in screening stringency.

Screening stringency has been recognized as one of the most important determinants of the growth of the disability insurance program. Previous studies have mainly analyzed the effects of known changes to screening stringency, affecting the eligibility criteria or the implementation of program rules, and have to a large extent concluded that such changes are important for program growth. To our knowledge this is the first paper that studies changes in screening stringency over time. Several studies have analyzed the effects of changes to formal eligibility criteria. Autor and Duggan (2003) use the 1984 liberalization of the disability determination process in the U.S. together with an unforeseen

---

4See, e.g., Bound and Burkhauser (1999) and Duggan and Imberman (2009).
increase in the effective replacement rate and declined demand for less skilled workers, and find a substantial impact of these forces on labor force participation of high school dropouts. Karlström, Palme and Svensson (2008) study the removal of looser eligibility criteria for workers aged 60 to 64 in Sweden in 1997, and find no effect on employment but a spillover effect to other social insurance programs. Staubli (2011) analyzes the effect of tightened eligibility criteria for men above a certain age in Austria in 1996. He finds a substantial decline in disability enrollment and an increase in employment, but also spillover effects into the unemployment and sickness insurance programs.

Other studies have exploited variation in the strictness of the eligibility screening process. Gruber and Kubik (1997) use the disability funding crisis in the U.S. in the late 1970s, which induced a sharp increase in initial denial rates. Exploiting variation across U.S. states, they find that stricter screening significantly increased labor force participation among older men. De Jong, Lindeboom and van der Klaauw (2011) use a large-scale experiment in the Netherlands in which caseworkers in certain regions were instructed to increase the screening of disability benefit applications. They find a reduction in long-term sickness absence and disability insurance applications from stricter screening, without negative spillover effects on the utilization of the unemployment insurance.

The paper proceeds as follows. Section 2 presents the theoretical model of the application decision to the disability insurance program. Section 3 discusses the empirical strategy and the data. Section 4 describes the Swedish disability insurance system and presents the estimation results. Section 5 concludes the paper.

2 The model
In this section we present a stylized model of an individual’s application decision to the disability insurance program. We focus on screening stringency as a key determinant of program inflow. In the model, an individual decides whether to apply for disability benefits or whether to continue working. If the individual applies and the individual’s health is lower than the disability benefit threshold, the individual receives disability benefits. If the application is rejected, the individual is assumed to return to employment. There is an
application cost which is increasing in the individual’s wage.

Let \( m_{it} \) be the disability benefit threshold that individual \( i \) faces at calendar time \( t \). This threshold can be expressed as

\[
m_{it} = m_t + \mu_{it},
\]

where \( m_t \) is the average disability benefit threshold at time \( t \) and \( \mu_{it} \) is an idiosyncratic error term, with zero mean, a cumulative distribution function \( F(\cdot) \) and a probability density function \( f(\cdot) \). The average disability benefit threshold, \( m_t \), is known by the individual.

When the individual applies for disability benefits, the caseworker can observe the individual’s health, \( h_{it} \). If the health is below the disability benefit threshold the individual faces, \( m_{it} \), then the individual is awarded disability benefits. Hence,

\[
D_{it}^{AC} = \begin{cases} 
1 & \text{if } h_{it} - m_t < \mu_{it} \\
0 & \text{if } h_{it} - m_t \geq \mu_{it},
\end{cases}
\]

where \( D_{it}^{AC} \) is an indicator of individual \( i \)'s application for disability benefits at time \( t \) being accepted.

The expected value from applying for disability benefits is given by

\[
V^{AP}(h_{it}, m_t, w_{it}) = p(h_{it}, m_t)V^{AC}(h_{it}, w_{it}) + [1 - p(h_{it}, m_t)]V^{RJ}(h_{it}, w_{it}) - C(w_{it})
\]

where \( V^{AC}(\cdot) \) is the value of the application for disability benefits being accepted, \( V^{RJ}(\cdot) \) is the value of the application being rejected, \( C(\cdot) \) is the application cost and

\[
p(h_{it}, m_t) = \Pr(h_{it} - m_t < \mu_{it}) = 1 - F(h_{it} - m_t),
\]

is the probability of the application being accepted. It follows that

\[
p_h(h_{it}, m_t) \equiv \frac{\partial p(h_{it}, m_t)}{\partial h_{it}} = -f(h_{it} - m_t) < 0
\]
and
\[ p_m(h_{it}, m_t) \equiv \frac{\partial p(h_{it}, m_t)}{\partial m_t} = f(h_{it} - m_t) > 0. \]

That is, the probability of obtaining disability benefits is decreasing in health and increasing in the average disability benefit threshold.

The value of an accepted application for disability benefits, \( V^{AC} \), is a function of the individual’s health, \( h_{it} \), and the wage, \( w_{it} \). The individual’s consumption, when receiving disability benefits, is given by \( \delta w_{it} \), where \( \delta \) is the compensation rate. The other source of utility is the individual’s health, \( h_{it} \). The value of receiving disability benefits is therefore increasing in all arguments.

The value of a rejected application for disability benefits, \( V^{RJ} \), is a function of the health, \( h_{it} \), and the wage, \( w_{it} \). Health generates utility and \( V^{RJ} \) is therefore increasing in \( h_{it} \). We assume that the individual returns to employment when being denied disability benefits. The value of being denied disability benefits is therefore increasing in the wage.

The application cost is increasing in the wage. This reflects the fact that in order to apply for disability benefits, for example in the U.S., the applicant has to leave his or her current employment and forego wage earnings. In Sweden, many individuals are on sick-leave when applying for disability benefits, which is more costly for high income earners. Hence,
\[ C_w(w_{it}) \equiv \frac{\partial C(w_{it})}{\partial w_{it}} > 0. \]

We can now determine when it is optimal for the individual to apply for disability benefits. The individual applies for disability benefits if the difference between the value of applying and the value of working is positive, i.e., if
\[ \Delta(h_{it}, m_t, w_{it}) \equiv V^{AP}(h_{it}, m_t, w_{it}) - V^{WO}(h_{it}, w_{it}) > 0, \]

where the value of working is increasing in both health and wage.

We assume that in the worst health state it is optimal for the individual to apply for
disability benefits while in the best health state it is optimal to work. Defining $h$ as the lowest and $\bar{h}$ as the highest possible health state, we can express these assumptions as

$$\Delta(h, m_t, w_{it}) > 0 \quad \text{and} \quad \Delta(\bar{h}, m_t, w_{it}) < 0.$$ (1)

In order to show that there exists only one level of health for which the individual is indifferent between applying for disability benefits and working, we are interested in how the difference between the value of applying and the value of working changes with health

$$\frac{\partial \Delta(h_{it}, m_t, w_{it})}{\partial h_{it}} = p_h(h_{it}, m_t) \left[ V^{AC}(h_{it}, w_{it}) - V^{RJ}(h_{it}, w_{it}) \right] + p(h_{it}, m_t) \left[ V^{AC}_h(h_{it}, w_{it}) - V^{RJ}_h(h_{it}, w_{it}) \right] + V^{RJ}_h(h_{it}, w_{it}) - V^{WO}_h(h_{it}, w_{it}).$$

In order to determine the sign of this derivative, we need to make two assumptions. The first one states that the value of an accepted application for disability benefits is larger than the value of a rejected application,

$$V^{AC}(h_{it}, w_{it}) > V^{RJ}(h_{it}, w_{it}).$$ (2)

The other assumption concerns the marginal utility of health. We assume that

$$V^{AC}_h(h_{it}, w_{it}) \leq V^{RJ}_h(h_{it}, w_{it}) \leq V^{WO}_h(h_{it}, w_{it}).$$ (3)

In other words, a marginal increase in health is least valuable for an individual receiving disability benefits and it is most valuable for a working individual.

These two assumptions together with the fact that $p_h(\cdot) < 0$ allow us to determine that

---

5This assumption is not very restrictive and can be motivated by a disutility of working which is decreasing in health. In other words, a healthy individual has a low disutility of working as well as a low probability of receiving disability benefits. This makes working more attractive than applying for disability benefits. An individual in bad health, on the other hand, faces a large disutility of working while at the same time having a high probability of applying successfully, which makes applying for disability benefits more attractive than working.
the difference between the value of applying and the value of working is decreasing in health, i.e.,
\[
\frac{\partial \Delta(h_{it}, m_{it}, w_{it})}{\partial h_{it}} < 0.
\]

Assumptions (1)–(3) guarantee the existence of a health level below which it is optimal to apply for disability benefits and above which it is optimal to continue working. Let us define this health cut-off level as \(h^{CO}(m_{it}, w_{it})\), where \(\Delta(h^{CO}(m_{it}, w_{it}), m_{it}, w_{it}) = 0\). The application decision can then be formalized as follows
\[
D^{AP}_{it} = \begin{cases} 
1 & \text{if } h_{it} < h^{CO}(m_{it}, w_{it}) \\
0 & \text{if } h_{it} \geq h^{CO}(m_{it}, w_{it}),
\end{cases}
\]
where \(D^{AP}_{it}\) is an indicator of individual \(i\) applying for disability benefits at time \(t\).

In order to analyze how the composition of the applicant pool changes when stringency changes, we need to know how this cut-off level changes with the average disability benefit threshold, \(m_{it}\). To determine this we compute how a change in \(m_{it}\) affects the difference between the value of applying and the value of working. We get that
\[
\frac{\partial \Delta(h_{it}, m_{it}, w_{it})}{\partial m_{it}} = p_{m}(h_{it}, m_{it}) \left[ V^{AC}(h_{it}, w_{it}) - V^{RJ}(h_{it}, w_{it}) \right] > 0,
\]
because of \(p_{m}(\cdot) > 0\) and assumption (2). In other words, the difference between the value of applying and the value of working is increasing in \(m_{it}\).

As illustrated in Figure 2, a reduction in stringency, i.e., an increase from \(m_{it}\) to \(m'_{it}\), pushes the difference between the value of applying for disability benefits and the value of working upwards, from \(\Delta(h_{it}, m_{it})\) to \(\Delta(h_{it}, m'_{it})\). This implies that the health cut-off level increases from \(h^{CO}(m_{it})\) to \(h^{CO}(m'_{it})\). We therefore know that
\[
\frac{\partial h^{CO}(m_{it}, w_{it})}{\partial m_{it}} > 0. \tag{4}
\]

As can be seen in Figure 2, the pool of disability benefit applicants increases from all individuals whose health falls in the interval \([h, h^{CO}(m_{it})]\) to all individuals with objective health in \([h, h^{CO}(m'_{it})]\). In other words, when stringency decreases, it becomes more at-
tractive to apply for disability benefits and individuals who were previously too healthy to apply find it now worthwhile doing so. The average health among applicants therefore increases when stringency decreases.\textsuperscript{6} However, since the sickest non-applicants are now applying for benefits, the average health among non-applicants increases as well when stringency decreases. How the relative health of applicants compared to non-applicants changes when stringency decreases is therefore not clear. In reality, however, applicants constitute at most 1.6\% of the population. Therefore, it can be argued that the sickest non-applicants becoming the healthiest applicants has virtually no effect on the large group of non-applicants while the inflow of those individuals to the relatively small group of applicants increases the average health of the applicants group. Similarly, we can argue that the fact that some applicants will be denied benefits and become non-beneficiaries has only a negligible effect and that therefore the relative health of beneficiaries compared to non-beneficiaries increases when stringency decreases.

In the above description, we focused on the relation between screening stringency and the health of disability benefit applicants. The result that we derived can be obtained also by changing other parameters of the model. The implications of this for the results in this paper are further discussed in Section 4.3.

\textsuperscript{6}This change resembles the results of Parsons (1991) who shows that the composition of disability benefit applicants depends, among other things, on the stringency of the screening process. However, he considers only two health classes, disabled and able.
3 Empirical strategy

3.1 Estimation

To study how the inflow to the disability insurance has been governed over time, we are interested in the relative health of new disability beneficiaries to non-beneficiaries. We analyze the ex-ante health of individuals entering the program, compared to individuals not entering the program in the same year, by using the ex-post mortality from all years in which the individuals are observed in the future. The strategy is employed using duration analysis. We assume that, under constant control of program inflow, the mortality of new disability beneficiaries at each age is proportionally related to the mortality of non-beneficiaries.\footnote{We could assume a specific form of the underlying distribution of longevity and perform the analysis within a full maximum likelihood framework. For a number of known longevity distributions this assumption is the same as assuming that there is a constant difference in the mean log longevity between the two groups. This will, e.g., be the case when longevity follows the Weibull and the exponential distribution. Thus the proportionality assumption allows us to estimate the mortality ratio without completely specifying the longevity distribution.}

We estimate the Cox proportional hazard model

\[ \lambda_{ic}(a) = g_c(a) \exp(\beta_tD_i(a)), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \]  

where \( \lambda_{ic}(a) \) is the mortality hazard of individual \( i \) in cohort \( c \) at age \( a \) and \( g_c(a) \) is the baseline hazard of cohort \( c \). \( D_i(a) \) is a step function that takes the value one after individual \( i \) begins receiving disability benefits at age \( a \) in year \( t \), else zero. Hence, \( t \) indicates the year in which the individual enters the disability insurance program. For an individual who is not receiving disability benefits during the studied time period, \( D_i(a) \) is zero for all ages. The estimated coefficient \( \exp(\beta_t) \) is the mortality ratio of new disability beneficiaries compared to non-beneficiaries in year \( t \). We have that \( \exp(\beta_t) > 1 \) if the mortality of disability insurance beneficiaries is larger than the general mortality at each age. Standard errors are estimated by clustering by birth cohort.

The cohort-specific baseline hazard captures changes in the mortality hazard across cohorts that may occur due to, for example, the expansion of education, changes in nutrition or advances within the health care system. Another reason to stratify the baseline
hazard by cohort is to solve the problem of length-bias sampling that comes from the fact that we observe different cohorts during different ages. Identification of $\beta_t$ comes from the within cohort comparison of mortality hazards at age $a$ for those who are awarded disability benefits at calendar time $t$ compared to those who have not yet obtained disability benefits. The individuals in the comparison group may, however, begin receiving disability benefits later if eligible then. Since individuals are no longer eligible for disability benefits above age 64, we censor the observations above this age.

The development of the mortality ratio over time indicates how the inflow to the disability insurance program has been governed. A constant mortality ratio suggests that the inflow has been governed in a similar way over time and that the variations in the inflow can be attributed to changes in the underlying health of the population. Large variations in the mortality ratio over time, on the other hand, suggest that the control of the inflow has varied. A high mortality ratio suggests a stricter control and a low mortality ratio suggests a looser control of the inflow to the disability insurance program.

Several of the reforms to formal eligibility criteria have concerned the age group 60–64. Differences in the inflow to the disability insurance program have also been noted between men and women and in different regions in Sweden. Therefore, we are also interested in the control of program inflow across these groups. We study how program inflow has been governed across age groups by estimating the model

$$\lambda_{ic}(a) = g_c(a) \exp(\beta_t D_t(a) + \delta^a_t D_t(a) I(a_i = a)), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \quad (6)$$

where the coefficients $\exp(\delta^a_t)$ are the mortality ratios of new beneficiaries to non-beneficiaries in age group $a$ at time $t$ compared to the mortality ratio in other age groups.

To analyze the development across gender and region of residence, we assume a proportional health at each age $a$ in the different populations $p$. We then have that

---

8The reason is that we do not know if an individual aged, say, 66 would still have remained in the control group. Had the eligible age been 66 years, his or her health could have been such that he/she instead would have received disability benefits if applying. Keeping individuals in the control group after the age of 64 could lead to an attenuation of the estimate. However, when we include data above age 64 in the analysis, the patterns of the results are similar.
where the coefficients $\exp(\delta_p^t)$ are the mortality ratios of new disability beneficiaries to non-beneficiaries in one population group at time $t$ compared to the mortality ratio in the other population group. The difference in mortality in general across the groups is captured by the parameter $\delta_0$. Regarding region of residence, we divide Sweden into the Stockholm region, the northern region Norrland, the middle region Svealand (excluding Stockholm) and the southern region Götaland. We use Stockholm as the benchmark and simultaneously estimate the mortality ratio of new beneficiaries compared to non-beneficiaries in Götaland, Svealand and Norrland compared to Stockholm by including main effects and interaction terms for each region except for Stockholm.

One potential criticism against the empirical strategy is that mortality is a bad measure of individual health. Mortality can be argued to capture a different margin of health than work capacity. Furthermore, the development of mortality over time is likely to reflect advances within the health care system rather than the general health of the population.

Another concern is that admittance to the disability insurance program may have a direct effect on health. This is not a concern for the results derived in the theoretical section, where health refers to the health status at the time of admittance. In the empirical strategy, however, we use ex-post mortality as a measure of ex-ante health and if there is a direct

\footnote{There are examples of using mortality as a measure of work capacity in the literature. E.g., Milligan and Wise (2012).}
effect of disability insurance recipiency on health, this may affect our estimates. The direction of such an effect is not clear, however. Kuhn, Wuellrich and Zweimüller (2010) find a significant increase in mortality from early retirement for male blue-collar workers in Austria, but no effect for females. Coe and Lindeboom (2008) find no detrimental effects of early retirement on health and, if anything, a slight short-term improvement. Moore (2012) finds positive health effects for a subgroup of DI beneficiaries.

3.2 Data
We use data covering all individuals in Sweden aged 30–64 from 1985 to 2008. The data originates from administrative records, collected and maintained by Statistics Sweden. For each individual, we observe birth year, gender and county of residence. We also have access to detailed spell data on the collection of disability benefits, provided by the Swedish Social Insurance Agency. For each year, we record if the individual receives disability benefits and whether it is the entry year to the program. Most individuals remain in the disability insurance program until retirement, once they enter. For individuals with multiple benefit spells, the first year of benefit collection is regarded as the entry year to the program. Since the data is based on disability benefit payments, there might be a slight delay of the first payment compared to the decision to award benefits. Finally, we add information about mortality, provided by the National Board of Health and Welfare. Mortality is measured until 2010 or until the year the individual turns 64.

The sample used for estimation consists of individuals aged 30–64 between 1985 and 2008 who are not already receiving disability benefits during the first year in which they are observed. This includes individuals who are observed from the beginning of the sample period, in 1985, and are not already receiving disability benefits at that time, as well as individuals entering the sample by turning 30 during 1986–2008 and not already receiving disability benefits at that age. The sampling strategy implies that different cohorts will be observed for different periods of time and at different ages. This problem of length bias sampling is solved by stratifying the baseline hazard by cohort in the Cox regression model. In total, the sample consists of about 100 million observations.

10 This simplification leads to only 2.4 percent of the registrations of program participation not corresponding to the data.
4 An application to Sweden

4.1 The Swedish disability insurance system

The Swedish disability insurance program replaces foregone earnings due to a lasting reduced working capacity for workers aged 19–64. The basis in the eligibility determination process is a clear relationship between medical causes and the reduction in working capacity. During certain time periods and for certain groups of workers consideration could also be given to the individual’s labor market situation. From 1972 to October 1991, individuals aged 60 to 64 could be granted disability benefits for pure labor market reasons, without a health assessment, if they were still unemployed when reaching the time limit in the unemployment insurance. From 1970 to January 1997, special rules for workers aged 60 to 64 further implied less strict health assessments, no test of employability and lower requirements for changing occupation or area of residence in order to find work. During the same time period, workers of all ages could be granted disability benefits for labor market reasons and health reasons combined, if they suffered from a reduced working capacity for medical reasons and had been unemployed for 1–2 years.

Since 1997, individual job opportunities should not be taken into account in the eligibility determination for disability benefits. In 2003, the disability insurance program was moved from the public pension system to the social insurance system as part of a public pension reform, but the formal eligibility criteria were unchanged. In 2005, the 21 regional offices of the Social Insurance Agency were integrated into one central authority. In the same year, the reassessment of eligibility for disability beneficiaries became stricter. In July 2008, the eligibility criteria to the program were substantially tightened.

Disability benefits can be granted part-time or full-time, depending on the extent of the work impairment. The compensation rate is 64 percent of the assumed income, up to a ceiling. The level of benefits has remained fairly constant over time. Before 2003, benefits were calculated according to the formula that applied to old-age pension. The average level of compensation was, however, similar before and after 2003. Disability benefits can be supplemented with payments from occupational insurances, covering the majority
of Swedish workers. Other social insurance programs include the sickness insurance, which replaces foregone earnings due to a temporarily reduced working capacity, and the unemployment insurance. Taking occupational insurances into account, disability benefits compare well with the level of sickness and unemployment benefits. The re-assessment of working capacity is also least strict in the disability insurance program, making it a desirable route for individuals who want to stop working permanently.

4.2 Results

Figure 3 and Figure 4 show the development of program participation in the Swedish disability insurance system along with the estimation results during 1986–2008. The dotted lines show the 95 percent confidence intervals of the estimates. The vertical dashed lines mark the reforms to formal eligibility criteria. The first vertical line indicates the removal of eligibility for pure labor market reasons for workers aged 60 to 64 in 1991. The second vertical line indicates the removal of the special eligibility criteria for workers aged 60 to 64, and the removal of eligibility for labor market and health reasons combined for workers of all ages, in 1997.

Figure 3(a) shows the entry rate to the disability insurance program in the full population aged 30–64, and Figure 3(b) shows the estimated mortality ratio of new disability beneficiaries to non-beneficiaries in the same group. The disability benefit entry rate has varied considerably during the studied period, and the overall picture from the estimation results is that substantial variation in program inflow remains after accounting for changes in the underlying health of the population. The mortality ratio in Figure 3(b) is larger than one, which confirms the expected result that new disability beneficiaries are in worse health than non-beneficiaries. The mortality ratio has varied substantially, however, with a mortality of awarded DI beneficiaries between 2.5 and 4.5 times higher than for non-beneficiaries during the period under study.

11The total compensation rate, including occupational insurances, is about 80 percent for workers in the private and local government sector and about 85 percent for workers in the central government sector, for earnings below the ceiling. Sjögren Lindquist and Wadensjö (2007) estimate that 96 percent of Swedish workers are covered by a collective agreement allowing for occupational insurance, and that almost all of these workers fulfill the criteria for receiving occupational insurance. Between 60 and 80 percent of workers claim occupational insurance when receiving disability benefits.
Figure 3(c) shows that DI entry has been higher in the age group 60–64 compared to 30–59. Figure 3(d) presents the estimated mortality hazard ratio of new disability beneficiaries compared to non-beneficiaries for the age group 60–64 compared to the age group 30–59. The relative mortality ratio below one suggests that beneficiaries in the age group 60–64 also have been in better health than beneficiaries in the age group 30–59, compared to non-beneficiaries of the same age. Likewise, Figure 3(e) shows that women have been more likely than men to be awarded disability benefits throughout the period. Figure 3(f) presents the estimated mortality hazard ratio of new disability beneficiaries compared to non-beneficiaries for women compared to men, and the relative mortality ratio below one also suggests that the health of new female beneficiaries typically has been better than that of male beneficiaries, compared to non-beneficiaries in the same gender group.\textsuperscript{12}

Figure 4(a) shows that the entry rate has been lowest in the Stockholm region and highest in Norrland throughout the period, whereas Svealand and Götaland have had intermediate entry rates. Figures 4(b), 4(c) and 4(d) present the estimated mortality hazard ratio of new disability beneficiaries compared to non-beneficiaries for Götaland, Svealand and Norrland compared to Stockholm. The relative mortality ratios below one are in accordance with the higher inflow to the program in these regions compared to Stockholm. This suggests that not only differences in the health of the population are behind the differential inflow across regions.

The development of the estimated mortality ratio over time indicates how the inflow to the disability insurance has been governed. The development of the disability entry rate over time corresponds well to changes in the estimated mortality ratios in general and across population groups. The rate of entry to the disability insurance program was stable at a high level during the late 1980s and early 1990s. The mortality hazard of new beneficiaries in the disability insurance program was also comparatively low, less than three times as large as that of non-beneficiaries. Workers aged 60 to 64 could be eligible for disability benefits for pure labor market reasons, which appears to have substantially

\textsuperscript{12}Recall that we control for the relatively lower mortality of women compared to men in the population by estimating the parameter $\delta_0$.\hfill
(a) Disability benefit entry rate, ages 30–64
(b) Estimated mortality ratio of new disability beneficiaries compared to non-beneficiaries, ages 30–64
(c) Disability benefit entry rate, ages 30–59 and 60–64
(d) Estimated mortality ratio of new disability beneficiaries compared to non-beneficiaries in ages 60–64 compared to ages 30–59
(e) Disability benefit entry rate for men and women, ages 30–64
(f) Estimated mortality ratio of new disability beneficiaries compared to non-beneficiaries for women compared to men, ages 30–64

Figure 3: Disability benefit entry rate and the estimation results, 1986–2008
impacted how program inflow was governed. Older workers were awarded benefits almost eight times as much as younger workers during the late 1980s, and the mortality ratio of older compared to younger new disability beneficiaries was at the lowest level during the period under study.

Figure 4: Disability benefit entry rate and the estimation results by region, 1986–2008

In October 1991, the eligibility to disability benefits for pure labor market reasons for workers aged 60 to 64 was removed. This was followed by a substantial drop in program entry in general. The decrease was mainly driven by the age group 60–64, for which the entry rate fell from between 6 and 7 percent between 1986 and 1992 to just above 2 percent in 1998. Women and men were similarly affected, and the drop in entry was apparent in all four regions of Sweden. The drop in program entry is well reflected in
an increase in the mortality ratio of new beneficiaries compared to non-beneficiaries. In particular, the mortality ratio of older compared to younger new beneficiaries increased dramatically following the reform, and remained at a higher level throughout the period.\textsuperscript{13}

Interestingly, the removal of pure labor market reasons for older workers in 1991 was followed by a convergence in the relative mortality ratio of new beneficiaries compared to non-beneficiaries across Swedish regions. During the late 1980s and the early 1990s, the mortality ratio of new beneficiaries compared to non-beneficiaries in G"otaland, Svealand and Norrland was significantly lower than in Stockholm. These regions also had a higher unemployment rate and worse labor market prospects on average than the Stockholm region. After 1993, the mortality ratio converged towards the Stockholm level, in particular for G"otaland and Svealand, and has since remained at a higher level. This suggests that the removed possibilities of taking labor market reasons into account affected eligibility determination and reduced the differences in the health of new beneficiaries across regions.

The next reform to formal eligibility criteria took place in 1997, when the special eligibility criteria for workers aged 60 to 64 and the eligibility for labor market and health reasons combined were removed. Interestingly, this reform does not seem to have had a lasting impact on the relative health of new disability beneficiaries compared to non-beneficiaries. There is substantial volatility in program entry for workers aged 60 to 64 around the reform that is well reflected in the relative mortality ratio of older compared to younger new disability beneficiaries. In 1997, there is a substantial drop in the mortality ratio of older compared to younger beneficiaries. This captures the anticipation effect of the reform that was found in Karlström, Palme and Svensson (2008). Applications under the old regime could be filed until 31 December 1996, which means that many of the award decisions were taken in 1997. The significant drop in the relative mortality ratio of new beneficiaries compared to non-beneficiaries across age groups in 1997 was followed by a significant spike in 1998, when the special rules had been removed. After

\textsuperscript{13}The transitory spike in entry in 1992 is not reflected in a change in how program inflow was governed. Anecdotal evidence suggests that a shortage of caseworkers led to a queue of cases during the preceding years, and that the spike in 1992 is due to a shifting of assessments across years. This would explain that the mortality ratio is unaffected.
that, however, the relative mortality ratio returned to the pre-reform level.

After the transitory responses to the 1997 reform followed a rapid increase in program entry into the disability insurance between 1998 and 2004. The large increase in program entry was accompanied by a steady decline in the mortality ratio of new beneficiaries compared to non-beneficiaries to a level comparable to that in the 1980s. The increase in the entry rate appeared both for younger and older workers and in all regions, and the relative mortality ratio of new beneficiaries compared to non-beneficiaries across age groups and regions was also fairly stable during several years. The parallel development of program entry for women and men, however, was broken, as program entry increased more rapidly for women than for men. The relative mortality ratio for women compared to men also decreased, suggesting that the large increase in disability benefit awards for women during this period may be explained by a looser screening of new female compared to male disability beneficiaries.

Finally, there has been a remarkable decline in the entry rate to the disability insurance program since 2004. The decline appeared across the board, for both age groups, both genders and in all regions. The rapid decrease in disability benefits awards from 2004 onwards is reflected in a marked increase in the mortality ratio of new beneficiaries compared to non-beneficiaries. Although no formal changes to the program were pursued until 2008, the mortality ratio increased already from 2005 to higher levels than ever before during the period under study. The mortality hazard of new disability beneficiaries in 2008 was more than four times higher than that of non-beneficiaries, suggesting that the access to the program was very restricted compared to previous levels. The subsequent decline in program entry since 2005 has again led to a convergence in entry rates across gender. The mortality of new disability beneficiaries compared to non-beneficiaries across gender also converged. In 2008 there is, for the first time during the period under study, no statistically significant difference in the mortality ratio of new disability beneficiaries compared to non-beneficiaries across gender.

The turning point coincides with the re-organization of the Social Insurance Agency into one central authority in 2005. OECD (2009) states that the main goal of restructuring the Social Insurance Agency was to strengthen central control and improve consistency
in the administration at the front line. The centralization allowed the Social Insurance Agency to increase uniformity of business processes. Applications were increasingly processed at the national centre, which allowed the frontline staff to engage with clients instead of processing benefit forms and medical certificates. The general view, according to OECD (2009), is that the fall in entry rates since 2005 was brought about by a gradual improvement in the way existing regulations were being implemented following the organizational changes.

4.3 Alternative explanations

In this paper we analyze how the inflow to the disability insurance program has been governed over time. In the theoretical model, we focused on the relation between screening stringency and the health of disability applicants. The result that we derived can be obtained also by changing other parameters of the model. Program generosity is a different instrument that policy makers can use to change the attractiveness of the disability insurance program and decreasing the benefits produces the same result as increasing screening stringency. Although benefit generosity may be important, the compensation rate has not substantially changed during the period under study.

Also factors related to the attractiveness of outside options may be important. In the model, the wage is constant and equal for all individuals, and there is no risk of unemployment. An increased value of working relative to collecting benefits, due to, for example, increased wages or improved employment prospects, could also lead to a decline in the relative health of disability beneficiaries compared to non-beneficiaries. The impact of labor market opportunities will be included in the estimates, and our measure can be taken as to capture the ability of the disability insurance program to withstand the influence of outside forces.

Figure 5 plots the estimated mortality ratio against the disability insurance entry rate and against several measures of the attractiveness of outside options. Figure 5(a) shows a clear negative correlation between the estimated mortality ratio of new beneficiaries compared to non-beneficiaries and the disability insurance entry rate. Figure 5(b) plots the estimate against the GDP growth rate, obtained from the National Institute of Economic
Figure 5: The estimated mortality hazard ratio of new disability beneficiaries to non-beneficiaries plotted against the DI entry rate and alternative explanatory factors.

Research, Figure 5(c) plots the estimate against the real wage growth for blue collar and white collar industry workers, obtained from the Swedish Mediation Office, Figure 5(d) plots the estimate against the replacement rate in the sickness insurance during 1993–2008 and, finally, Figure 5(e) plots the estimate against the average replacement rate in
the unemployment insurance during 1993–2008. If the attractiveness of outside options would be important for explaining the inflow to the disability insurance program over time, we would expect a positive relationship between these measures and our estimate. As can be seen from Figure 5, however, there is no clear relationship to our estimate for any of these measures. This suggests that screening stringency is a major component of our estimate.

5 Conclusion

Changes in the screening stringency of applications to the disability insurance program are potentially important for explaining program growth. Screening stringency is, however, inherently difficult to observe since it depends on the implementation of program rules as well as formal eligibility criteria. In this paper we present a theoretical model of the application decision to the disability insurance program. We further propose an empirical strategy for capturing how the inflow to the disability insurance program has been governed over time. We use a Cox proportional hazard model in which we compare the mortality of new disability beneficiaries to non-beneficiaries, taking age into account and allowing for cohort-specific mortality hazards.

We study the inflow to the disability insurance program in Sweden between 1986 and 2008. Overall, our results suggest that substantial variation in program inflow remains after accounting for changes in the underlying health of the population. We find that the estimated changes in the mortality ratio of new beneficiaries compared to non-beneficiaries correspond well to the development of the disability benefit entry rate, overall as well as within groups defined by age and gender. Changes in formal eligibility criteria are well reflected in the mortality ratio. The removal of eligibility for pure labor market reasons for workers aged 60–64 in 1991 can be seen in a substantial and lasting increase in the relative mortality ratio of older compared to younger beneficiaries. Also the removal of the special elderly rules and eligibility for labor market and medical reasons combined in 1997 is reflected in the relative health of new beneficiaries across age groups, although this reform seems to have had more of a transitory impact. During the period under study,
program inflow was more strictly governed in Stockholm compared to other regions, but the relative mortality ratios across regions have converged after the possibilities of taking labor market reasons into account were removed in the 1990s.

We also find that fluctuations in the entry rate during periods of no formal program changes are reflected in changes in the relative health of new disability beneficiaries. The large increase in program entry during the early 2000s is related to a decrease in the relative mortality of new beneficiaries, and the large inflow of women compared to men during this period corresponds to a relatively lower mortality ratio of female compared to male beneficiaries. Most strikingly, the rapid fall in program entry since 2005 is reflected in a substantial increase in the mortality ratio of new beneficiaries compared to non-beneficiaries, although no formal changes to the program were being pursued. The turning point in 2005 coincides with the re-organization of the regional Social Insurance Agency offices into one central authority.
References


Publication series published by IFAQ – latest issues

Rapporter/Reports

2013:1 Olsson Martin "Anställningsskydd och föräldrarelaterad frånvaro"
2013:2 Angelov Nikolay, Per Johansson and Erica Lindahl "Det envisa könsgapet i inkomster och löner – Hur mycket kan förklaras av skillnader i familjeansvar?"
2013:3 Vikman Ulrika "Så påverkar föräldraförsäkringen nyanlända invandrare etablering på arbetsmarknaden"
2013:5 Eliasson Tove "Löneutveckling bland invandrade och infödda – betydelsen av arbetsplatser och yrken"
2013:6 Katz Katarina and Torun Österberg "Unga invandrade – utbildning, löner och utbildningsavkastning"
2013:7 Angelov Nikolay, Per Johansson and Erica Lindahl "Kvinnors större föräldraansvar och högre sjukfrånvaro"
2013:8 Johansson Per, Lisa Laun and Tobias Laun "Hälsan hos nybeviljade förtidspensionärer över tid"

Working papers

2013:1 Nekby Lena, Peter Skogman Thoursie and Lars Vahtrik "Examination behavior – Gender differences in preferences?"
2013:2 Olsson Martin "Employment protection and parental child care"
2013:3 Angelov Nikolay, Per Johansson and Erica Lindahl "Is the persistent gender gap in income and wages due to unequal family responsibilities?"
2013:4 Vikman Ulrika "Paid parental leave to immigrants: An obstacle to labor market entrance?"
2013:5 Pingel Ronnie and Ingeborg Waernbaum "Effects of correlated covariates on the efficiency of matching and inverse probability weighting estimators for causal inference"
2013:6 Forslund Anders, Linus Liljeberg and Leah von Trott zu Solz "Job practice: an evaluation and a comparison with vocational labour market training programmes"
2013:7 Eliasson Tove "Decomposing immigrant wage assimilation – the role of workplaces and occupations"
2013:8 Katz Katarina and Torun Österberg "Wages of childhood immigrants in Sweden – education, returns to education and overeducation"
2013:9 Angelov Nikolay, Per Johansson and Erica Lindahl "Gender differences in sickness absence and the gender division of family responsibilities"
2013:10 Johansson Per, Lisa Laun and Tobias Laun "Screening stringency in the disability insurance program"

Dissertation series

2012:1 Laun Lisa "Studies on social insurance, income taxation and labor supply"