Essays on Labor Supply and Health

Michael Jørgensen

Ph.D. Thesis

September 2012

Department of Economics, Copenhagen University

and

SFI – The Danish National Centre for Social Research

Contents

Acknowledgments

This thesis is the result of my Ph.D. studies at the Department of Economics, Copenhagen University (KU) and SFI – The Danish National Centre for Social Research (SFI). Funding was provided by the Danish Social Sciences Research Council (FSE-09-063848) and SFI. I benefited greatly from the support of SFI and KU, which provided me with the financial means to travel to conferences and courses, with data for my projects, and with a great working environment.

I thank my supervisors Søren Leth-Petersen and Paul Bingley who have been extraordinarily committed and have supported me greatly with their invaluable advice and guidance. Their ever positive and constructive comments have improved my work.

I thank my colleagues at SFI and KU, my fellow Ph.D. students at both institutions, my "office" mates at "the 3th floor". Furthermore, I was lucky to work with Natalie Reid and Ellen Rosenfield (Berkeley) while writing this thesis. Thanks for teaching me invaluable lessons on academic writing.

I am grateful to have participated at the NBER International Social Security group Munich 2012 and Rome 2012 workshops and for comments received in these workshops.

And finally, I would like to thank Signe Hastrup Poulsen for providing invaluable help with proofreading and setup of the Ph.D.-dissertation, and in particular for being supportive and encouraging through the whole process.

Michael Jørgensen
Copenhagen, August 2012

Summary

This dissertation is comprised of three self-contained chapters in applied micro-econometrics. All three chapters use Danish register data to answer empirical questions within the topics of labor supply and health. Following are the three abstracts that summarize each chapter.

Health, Disability Insurance, and Retirement in Denmark

The aim of this paper is twofold. First to measure the effects of pension program provisions on retirement age controlling for health based on register and SHARE health data and secondly to assess the importance of the disability insurance program for retirement controlling for health. We explore to what extent early retirement is determined by the provisions of disability insurance, other pension programs and health status. A 1999 pension reform rolled out through 2006 brought the largest changes to Danish retirement incentives in 20 years. Early retirement for the insured was made less generous, old age pension age was reduced, and disability insurance awards became more stringent. We characterize each of these changes using an option value model for the most important routes to retirement using administrative data on the population of single person households aged 58-67 during the years 1996-2008. Controlling for health (length of hospitalization, diagnosis severity, sickness absence from work), greater stringency in awarding disability insurance has modest effects on retirement, whereas reforms to incentives in an early retirement program had large effects on retirement age, probably due to the relative magnitudes of the reforms. We find that pension program incentives in general are important determinants of retirement age, and individuals in worse health retire younger and are less responsive to incentives.

How Does a Health Shock to Self or Partner Affect Economic Incentives to Retire?

This paper exploits rich administrative data to analyze how a health shock to one-self or one's partner affects the marginal economic incentive to retire. One important challenge when trying to get unbiased estimates of the effect of health on retirement is the endogenous character of the relationship between health and retirement. To circumvent this endogeneity, the paper exploits a 1999 reform of an early retirement program and the arguably unanticipated timing of a health shock. The paper contributes to the existing literature on retirement and health in several ways. First it models the interaction between a health shock and economic incentives to retire, thereby allowing health to affect the response to economic incentives. Second, the paper exploits the unanticipated timing of a health shock, not only levels of health, which helps circumvent the endogeneity problems that health measured in levels gives

rise to. Third, it uses a reform of an early retirement program that exogenously altered economic incentives without altering preferences or health in the short run. Fourth, it uses objective diagnoses from a large register sample, which allows us to analyze the effect of health shocks instead of health levels, which, due to small sample properties, are often not possible to analyze from survey samples. The main conclusions are that economic incentives have strong effects on retirement for both men and women, but that those who receive a health shock are much less responsive to economic incentives, regardless of whether the health shock strikes them or their partners. A comparison of men's and women's responses finds no differences in men and women's marginal responses to economic incentives due to a health shock to their partners–only when the health shock hits them self, then women react much stronger than men in their reduction to the economic incentive. These final results indicate that complementarity in leisure dominates substitutability when a health shock hits a partner.

Can we reform Disability Insurance and increase labor supply?

Disability Insurance (DI) benefits may distort labor supply decisions for those capable of working because defining and observing the true health of an applicant is difficult. This difficulty might encourage moral hazard on the applicant's part. Moral hazard problems may be reduced directly through stricter screening of applicants or by reducing over-all DI benefits. We investigate the effects of attempts to mitigate moral hazard in the DI program by raising the stringency of the screening process for applicants. We measure the effect on DI awards and labor supply of two DI reforms that took place in Denmark during the 1990s: (1) state audits of local awards, (2) reduction in state rebates to local authorities for DI awards to applicants aged 60-66. Benefit generosity at the individual level remained unchanged throughout the estimation period, keeping economic incentive effects stable. State audits of local county boards showed that almost no counties awarded too few DI pensions according to the state audit board, whereas many awarded too many, and when these high awarding counties were audited, they increased rejection rates the following year by on average 21%. Reduced state rebates to municipalities for DI awards for those aged 60-66 increased rejection rates by 5 percentage points for this age group compared to those aged < 60. Instrumental variables estimates show that a 10% rise in rejection rates leads to a 6.4% increase in labor supply.

Health, Disability Insurance, and Retirement in Denmark

Paul Bingley, SFI – The Danish National Centre for Social Research
Nabanita Datta Gupta, Department of Economics and Business, Aarhus University
Michael Jørgensen, SFI – The Danish National Centre for Social Research
Peder J. Pedersen, Department of Economics and Business, Aarhus University

Abstract

The aim of this paper is twofold. First to measure the effects of pension program provisions on retirement age controlling for health based on register and SHARE health data and secondly to assess the importance of the disability insurance program for retirement controlling for health. We explore to what extent early retirement is determined by the provisions of disability insurance, other pension programs and health status. A 1999 pension reform rolled out through 2006 brought the largest changes to Danish retirement incentives in 20 years. Early retirement for the insured was made less generous, old age pension age was reduced, and disability insurance awards became more stringent. We characterize each of these changes using an option value model for the most important routes to retirement using administrative data on the population of single person households aged 58-67 during the years 1996-2008. Controlling for health (length of hospitalization, diagnosis severity, sickness absence from work), greater stringency in awarding disability insurance has modest effects on retirement, whereas reforms to incentives in an early retirement program had large effects on retirement age, probably due to the relative magnitudes of the reforms. We find that pension program incentives in general are important determinants of retirement age, and individuals in worse health retire younger and are less responsive to incentives.

1. Introduction

Labor force participation of older persons has varied greatly both between countries and within countries over time. Individual health status, labor market conditions and social security program provisions have all played a role in these changes. At the interface between social security, labor market conditions and health, disability insurance (DI) programs may also play an important role for many persons as they move from employment to retirement from the labor market. In principle, it may be the case that changes in DI participation rates reflect changing health and changing labor market conditions. However, trends in DI participation appear to be unrelated to changes in mortality and health. In many countries DI effectively provides early retirement benefits before eligibility for other social security programs begins. If this were not the case, differences in health between countries would need to be much larger than those revealed in comparable survey data (Milligan and Wise, 2011). This begs the question: given health status, to what extent are the differences in labor force participation across countries determined by the provisions of disability insurance programs?

Social security programs have been shown to provide strong incentives for older workers to leave the labor market at certain ages (see e.g. Gruber and Wise, 2004; Bingley et al., 2004). In order to isolate the retirement incentive derived from DI programs controlling for health it is important to characterize the full retirement incentive of the social security systems and private pensions, because competing pathways to retirement are important for understanding the use of the DI program. The option value model (OV model) proposed by Stock and Wise (1990a) is very useful for this purpose, because it converts all future income streams from different pathways to retirement into one single economic incentive, the option value of retiring today compared to retiring at all future dates.

It is important to control for health, in order to identify the effects of DI provisions on early retirement. The policy-relevant thought experiment is: for a given health status, how do different aspects of a DI program affect the labor force participation of older workers? Answering this question is a challenge because measuring health is notoriously difficult.[1] Health conditions can be controlled for by using comparable health measures for populations faced with different DI programs.

---

[1] Self-reported health suffers from recall error, reference group issues and justification bias. Administrative data measures use of health care services, rather than health per se. Mortality is comparable, but occurs with long and indeterminable lags from program exposure.

To control properly for health one either needs repeated survey questions or administrative data on health observations spanning DI reforms within countries, or needs to ask similar questions in different countries with different DI programs. We control for health conditions by using a wide range of health questions from SHARE survey data that are comparable across countries, and thus also with other studies. Two disadvantages to using SHARE health information is that health is self-reported and that SHARE-DK only has a modest number of observations. Therefore we also use a large administrative register sample with information on objective health measures over several years to estimate the relationship between health and retirement.

We find that pension program incentives in general are important determinants of retirement age, and individuals in poor health are significantly less responsive to economic incentives. We also find that the level of stringency in the DI program only has small effects on labor supply in Denmark for individuals aged 59-66. This is relative to the large incentive effects from an early retirement program without health-related eligibility criteria.

The remainder of the paper is organized as follows. Section 2 describes pension programs, reforms, and pathways to retirement. Section 3 presents an OV model of retirement controlling for health, and illustrates OV calculations for different routes to retirement. Section 4 describes our data in general, defines the health measures we use from registers and SHARE and estimates the probability of being awarded DI conditioning on DI stringency. Section 5 presents retirement model estimates, observed and predicted hazard rates, and counterfactual simulations. Section 6 concludes.

1.1 Literature review of methods

There are essentially two approaches to modeling individual retirement behavior which consider the future economic consequences of current work decisions. Both are approaches to solving dynamic programming models. One approach, exemplified by Rust (1989), solves the dynamic programming problem explicitly by computing expectations of the maximum of the values of choices in future periods. The other approach, exemplified by Manski (1988), approximates the exact solution by computing the maximum of expected values of choices next period. The approximation is computationally convenient, but it is an empirical question as to whether the approximation is reasonable. In the simplest context of a single agent optimal stopping retirement model, Lumsdaine, Stock and Wise (1991) have found both approaches to have similar predictive power on real world data, whereas Stern (1999) finds the approximation to be inferior on simulated data. While these comparisons were inconclusive, the different approaches each have merits for different extensions of this simple retirement model.

In this paper we aim to model retirement age as a function of pension incentives and health, where disability insurance programs play an important intermediary role. Explicit computation of expectations of maximum values would require modeling expected future health transitions, DI applications and awards as well as retirement decisions. This would require simplifying assumptions and approximations for the explicit solution approach. Here we follow the maximum of expected values approach of option values invented by Stock and Wise (1990) and embed DI awards and health in the expectations calculations.

1.2 Literature review on substance

Both financial incentives and health are important determinants of the retirement decision (Hurd, 1990 and Lumsdaine and Mitchell, 1999). The magnitude of the importance of financial incentives is best documented in the comparative cross-country micro study by Gruber and Wise (2004). There is more controversy about the size of health effects because of the difficulties of measuring health and the potential simultaneity between health and work, as described in Lindeboom (2005).

Modeling incentive and health effects on retirement is complicated in the US context by health expenditure uncertainty linked to employer-provided health insurance (see Blau and Gilleskie, 2008). This leaves three papers most relevant to our research agenda of modeling joint incentives and health effects on retirement. Lindeboom and Kerkhofs (2008) estimate a joint model of retirement, health and health measurement error. Their contribution is to deal with both the simultaneity and measurement issues of health reporting in a retirement model. However, financial incentives are not dynamic, but are rather characterized as static replacement rates. Banks et al. (2007) are the first to model the interaction between dynamic retirement incentives and health. Incentives are calculated as peak values which are a simplification of option values where individuals compare future income streams ignoring work disutility and risk aversion. Nevertheless, they find that financial incentives are somewhat more important for those without self-assessed health problems. Erdogan-Ciftci et al. (2011) model interaction between dynamic retirement incentives and health levels and changes. They also use a peak value simplification and similarly find that those in good initial health react most to financial incentives.

Our contribution is first to go further than Banks et al. (2007) and Erdogan-Ciftci et al. (2011) by using an option value approach to better characterize the future value of current decisions, by allowing for disutility of work and risk aversion. Note that this contribution is in common with companion papers in the edited volume to which this paper contributes. Second, while we do not have an explicit model for measurement error as does Lindeboom and Kerkhofs (2008), we have access to two independent sources of health information, from a survey and administrative registers, and we create a common health index and compare results across measures. Note that this contribution is unique to our current paper.

2. Pathways to retirement

There are three main pension programs supporting income in retirement: 1. A contribution-based but largely tax financed Post Employment Wage program (*efterløn*, hereafter PEW) which is essentially unemployment insurance benefit without a job search requirement available for ages 60-66 and after 2006 from 60-64. 2. A disability insurance program (*førtidspension*, hereafter DI) is available for those aged 18-66 and later 18-64 who have permanent social and health impairments that reduce work capacity. 3. Old Age Pension (*folkepension*, hereafter OAP) is a demogrant available from age 67 and after 2006 from age 65, based on years of residence. Our period of analysis 1995-2008 is chosen to span reforms in DI stringency and PEW/OAP incentives in order to provide variation by which to identify the effects of program provisions on the retirement age for older workers.

2.1 Post-Employment Wage and Old Age Pension Programs

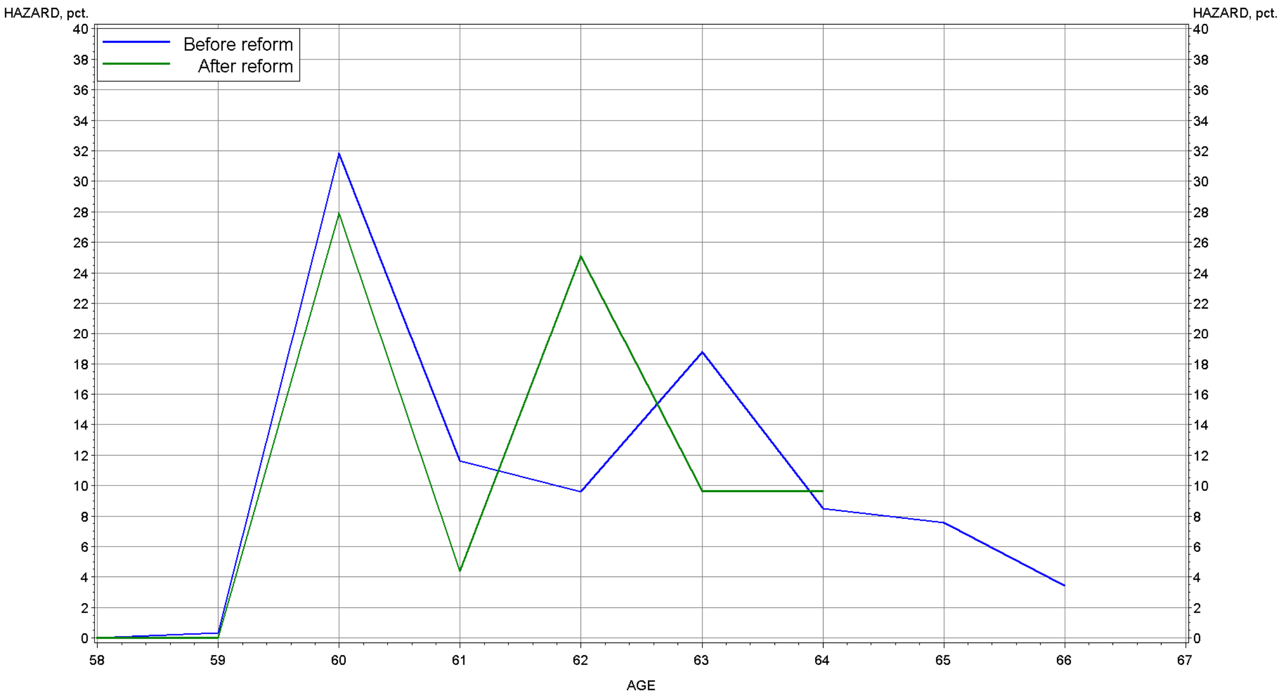
In this paragraph we describe the main elements of each program and the changes that occurred during this period. PEW was introduced in 1979 for ages 60-66 and existed largely unchanged until reforms in 1992 and 1999. The 1992 rules are relevant for the first part of our sample period. Eligibility from 1992 to 1999 required membership in an unemployment insurance fund for at least 20 of the last 25 years. An individual was allowed to work for a maximum of 200 hours. If the 200 hours was exceeded, it resulted in a permanent disqualification from the program. For individuals claiming PEW at ages 60-62, the benefits for the first two years were at the level of unemployment insurance and reduced to 80% for the last 4 years. Delaying PEW until age 63 or older gave benefits at 100% of the UIB level until age 66. This policy obviously incentivized retiring at age 63 rather than at younger or older ages. In 1995 unemployment insurance (UI) fund membership history requirements were increased from 20 to 25 out of the previous 30 years. Until 1999 only payouts from life annuities in occupational pensions were means tested.

The PEW reform was announced in March 1999 and enacted by the 1 of July 1999. Means testing of payouts from all contributory pensions–whether they were paid out or not–was introduced for those claiming PEW at age 60 and 61. Those eligible for PEW and not retiring now accumulate a quarterly €1.600 bonus beginning age 62. This reform shifted the retirement age incentive spike from age 63 down to age 62. The previous limitation of working at most 200 hours per year was removed. Working while claiming benefits was instead made subject to a high effective marginal tax rate. UI fund membership history requirements were further increased to 30 out of the last 35 years. Contributions were unbundled from UI and became separately elective.

An important element of the 1999 reform was the reduction in OAP age from 67 to 65. Those aged 60 and above at enactment (born before July 1. 1939 were unaffected and could first claim OAP at age 67, whereas those born later could claim OAP from age 65. The change in OAP age was implemented from July 2004 through June 2006 and the maximum age for claiming PEW benefits changed accordingly.[2]

Figure 1 shows the hazard rate to PEW before and after the reform in 1999. Before the reform the incentive was to wait until age 63, and after the reform the incentive was to wait until age 62–this clearly shows in the retirement pattern in figure 1. Both before and after the reform there was a huge increase in retirement at age 60 when PEW became available and again at age 62 (after the reform) and age 63 (before the reform).

Figure 1: Retirement hazard rate to PEW before and after reform in 1999, 1992-2009



As explained earlier, the OAP age was reduced from age 67 to age 65 for people who became 60 after July 1, 1999. Therefore it was possible to retire to PEW until age 66 before the reform.

[2] This was also the reason why the incentive to retire in the PEW program was reduced from age 63 to age 62.

2.2 Disability Insurance

Disability insurance existed in essentially the same economic form in the period 1984-2002 but with some stringency tightening in the 1990's. It was available to those with permanent social or physical work impairments depending on three levels of severity/generosity. During this period, real benefit levels were not changed but several stringency measures were introduced at different times. It is useful to distinguish between three separate stringency reforms:

1. During 1995-2002 a series of selective municipal award audits were undertaken. Each year two out of Denmark's fifteen counties were chosen and a random sample of 150 new awards from different municipalities in each chosen county was audited.
2. In 1997, central government rebates to municipalities were reduced for DI expenditure on DI to individuals aged 60 and above. Previously, the central government had refunded all DI expenses for this age group, but post-reform only 50% was refunded, bringing refunds into line with those for younger age groups.
3. In 1998, a filtering reform was introduced, which required municipalities to first consider whether other locally administered programs might be relevant before processing an application for DI. Important alternatives were work re-habilitation and disability wage subsidies (*fleksjob*).[3] Disability wage subsidies were by far the most common alternative route taken following the filtering reform.

In 2003, the government simplified DI by reducing the number of levels from three to one but also introduced an array of condition-and needs-specific financial additions. These additions make net changes to incentives due to the reform difficult to untangle.

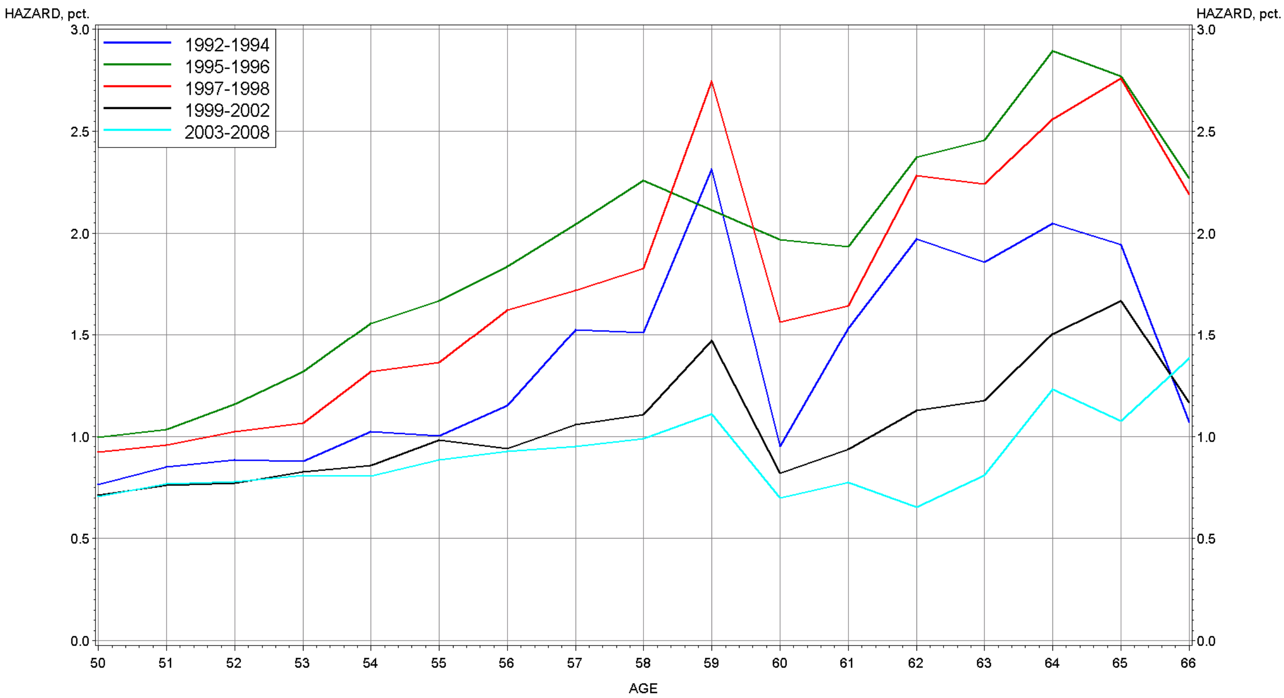
Relevant related programs for those in short-term poor health, with short-term or permanent work impairments but some remaining work capacity are sickness benefits (*sygedagpenge*), rehabilitation benefits (*revalidering*) and disability wage subsidies (*fleksjobs*) respectively. None of these programs include retirement, but they are worth mentioning because of their relevance at the interface between health, work and retirement. Work sickness absence benefits and rehabilitation are awarded temporarily. Disability wage subsidies are a payment at the level of the minimum wage for permanently

[3] Statistically, participants in both alternative programs are still in the labor force. The former program has temporary eligibility whereas the latter is permanent.

reduced work capacity. Individuals in this program are employed, or unemployed and seeking work, and therefore not retired for modeling purposes. Before receiving their benefits, they belong to the pool of DI applicants, which thereby influences DI award stringency.

Figure 2 illustrates retirement to DI pension from age 50 to 66 split into the above mentioned reform periods. The first point to notice is the relatively much lower DI hazard rate compared to the PEW hazard rate. The main reason for this is the different windows where the two programs are available (DI is available from age 16-64/66 while PEW is only available from age 60-64/66). Secondly there is an increase in the hazard rate depending on age although it declines over time. This increase is probably due to the normal deterioration of health with age.

Figure 2: Retirement hazard rate to DI split by reform periods, 1992-2009



There are several possible explanations for the decline in the hazard rate over time: a) the reduction in the rebate to municipalities for DI awards for people aged 60+ after 1996, b) the filtering reform from 1998 which required that all other possibilities be exhausted before a person was allowed to apply for DI[4], and c) the improvement in the business cyclical which took place from the mid 1990's and until 2008.

[4] Some applicants who prior to 1998 would have been awarded DI pension would now be awarded Disability wage subsidy.

From figure 1 and figure 2 it is clear that almost all retirement from age 60 was to PEW while retirement to DI only accounted for a smaller part. This becomes even more clear, if we compare figure 1 with figure 1a in the Appendix A that illustrates retirement to all states. Based on figure 2 we do not expect DI provisions to have a large influence on retirement in Denmark.

3. The Option Value model

The OV model compares the difference between the present discounted value of all future income streams from retiring today and retiring at the age that gives the maximum present value of all future income streams.

Stock and Wise (1990a) formulated the OV model as an approximation to the explicit solution of the Dynamic Programming model (DP model). They formulated the OV model as a deterministic model–you get what you expect–in order to reduce the complexity of the DP model, where uncertainty is handled by incorporating multidimensional integrals, which makes the DP model more computationally demanding.[5] In the Danish settings uncertainty is a much smaller problem than in many other countries, because health care is free and the welfare system ensures a relatively high income when you retire or lose your job. Because of this we argue that the financial shock that a health shock induces is of less importance in Denmark, because of the universal health care coverage in the country.[6]

The OV model emphasizes the economic incentives that come from the interaction of public and private pension programs, which is very important in order to explain retirement behavior in Denmark, where most of the retirement incentive comes from the interaction between public and private pensions.

Appendix B describes the OV model in detail.

3.1 Option value calculations by pathway

Figure 3 shows the OV for PEW pathway before and after the reform in 1999. As expected, the biggest average OV is achieved if workers stay at their jobs until OAP age, because your income stream from

---

[5] Belloni (2008) show that both the Dynamic Programming model and the OV model are dynamic, updating, and forward looking. The main difference between the two models is in the maximization of the utility function and the treatment of uncertainty. Theoretically the DP model is preferable because it uses a more formal solution to the intertemporal utility maximization problem. Empirically the two type of models have proven to do equally well in predicting in- and out-of sample (see e.g. Stock and Wise, 1990a; Lumsdaine et al., 1991; Burkhauser et al., 2003).

[6] In Rust and Phelan (1997) uncertainty about future health is very important because health care is not universal and coverage is very low for certain groups. If you retire early you might lose your coverage or change to a lower coverage, which makes uncertainty about future health much more important.

work and private pensions will be significantly higher than benefit streams from PEW with means testing of private pensions.[7]

The incentives in the PEW program are also visible in figure 3. Before 1999 there was a large drop in the OV from age 59 to age 60 and no change in the OV from age 60 to age 61, indicating that the utility gain for waiting to retire until age 60 is large, whereas there is no gain for waiting from age 60 to age 61. The same pattern is seen at age 63 before the reform and at age 62 after the reform. We therefore conclude that the OV is capable of capturing the economic incentives of the PEW program.

Figure 3: Option value before and after the reform in 1999 by age, 1995-2008, PEW

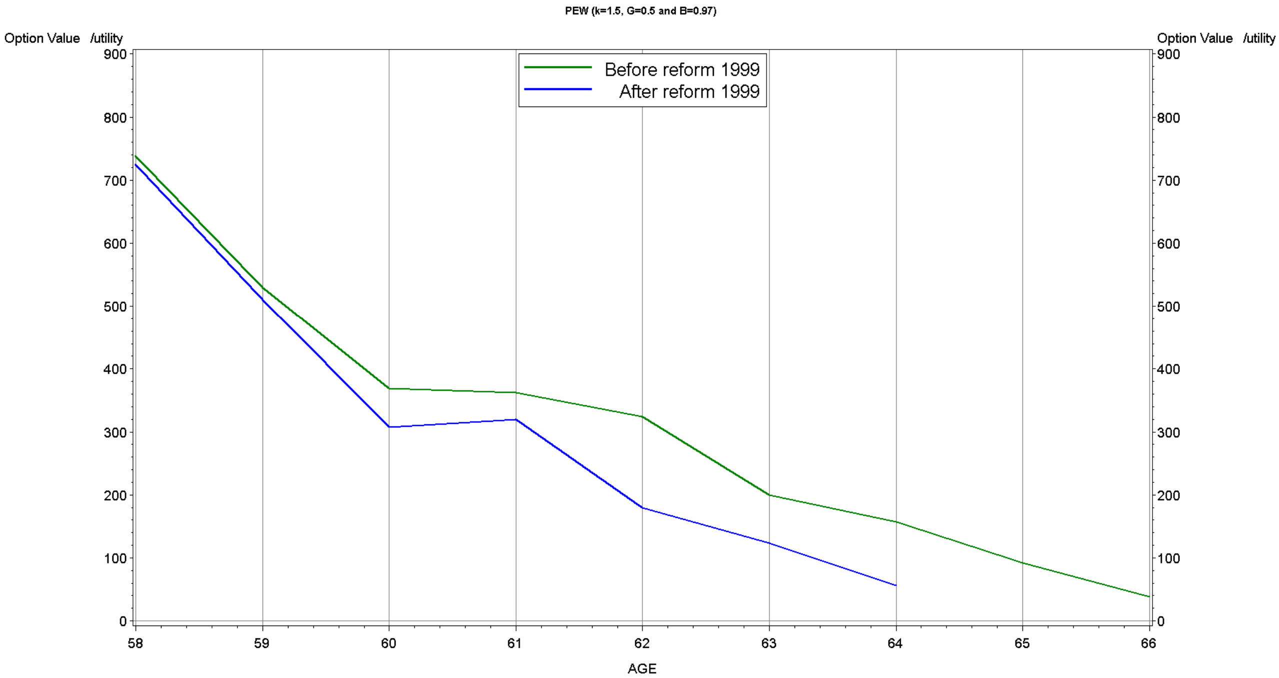


Figure 4 compares the OV for the DI pathway and the PEW pathway after the reform in 1999. Those eligible for PEW have higher average OV's at all ages. The main reason for this is that this population has a stronger labor market attachment and is better educated, resulting in higher earnings profiles. From figure 4 we also see that the OV is almost linear in age for the DI pathway, while this is not the case for the PEW pathway. This indicates that the incentives in the PEW program are much stronger at different ages, while it is almost unaffected by age in the DI program.

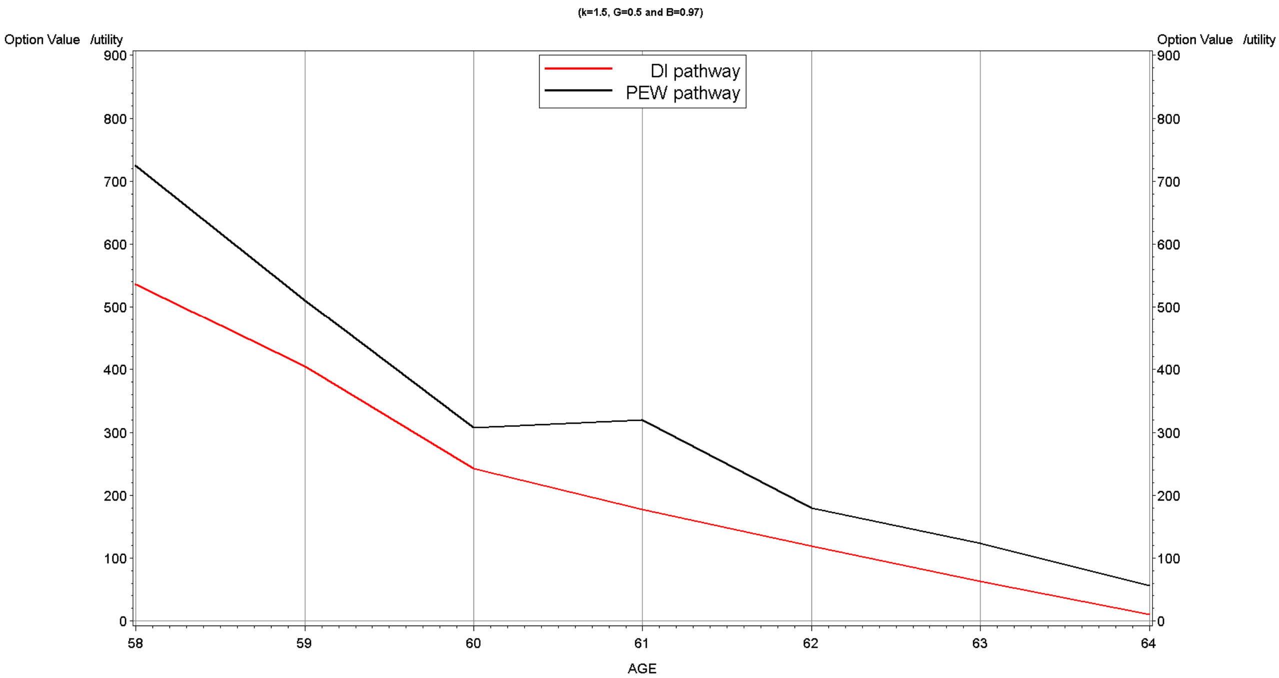
---

[7] This is because that for each extra year you work, you save one extra year for your pensions and they also have to be paid out one year less. Besides that, you earn interest on the entire pension wealth for one more year, which is very important too.

Figure 2a in Appendix A shows the OV before and after the reform of the PEW in 1999 for the DI pathway. As expected, the OV is almost completely linear with respect to age, because there is no built-in incentive in the DI program to retire at a certain age, and it does not change much after the 1999 PEW reform, which indicates a low substitution rate between the two programs. The OV measurement is higher at every age in the period 1995-1999 than in the period 2000-2008. There are two possible explanations for this:

a) In the 1990's occupational labor market were expanded to cover about 85% of the workforce, whereas they only covered about 30% in the early 1990's.[8] This difference in coverage resulted in a higher income stream in the post reform period,
b) In 2003, the benefit level increased on average with about 20%, which made it relatively more attractive to receive DI in the post reform period.

Figure 4: Option value for the DI and PEW pathway after the 1999 PEW reform, 2000-2008



[8] A lot of occupational pensions contain health insurance that is paid out as a life annuity pension, when you start receiving DI pension.

4. Data description and health measures

Since the aim of this paper is both to measure the effects of pension program provisions on retirement age, and compare data across countries, we use two sources of data. We use a 50% randomly drawn register based sample of the population to calculate the OV and register health measures, and we use a sample from SHARE to calculate a health index (using a principal component analysis) based on multiple health questions available to all countries in the study. We use the 2004 and 2006 waves of SHARE to calculate the health index, and we use register data for the period 1995 to 2008.

4.1 Definition and descriptive statistics of the estimation samples

Table 1 shows the distribution of selected variables for the registers and SHARE respectively. There are 265.273 individuals in the register data and 376 of them can be found in the two SHARE waves.

Table 1: Descriptive statistics, registers vs. SHARE, age 59, estimation samples

	Registers		SHARE	
	Men	**Women**	**Men**	**Women**
# Individuals	142,314	122,959	212	172
Total assets	195,149	160,056	294,111	154,439
Income at age 58 (1.000 DDK)	59	43	61	41
Days in hospital	1.54	1.22	0.89	1.36
Days sick from work	3.16	4.13	1.58	4.35
Reduction survival 5 years – BETA	-0.79	-0.67	-0.55	-0.84
Schooling				
Basic school	27.40	34.73	25.00	39.53
General upper secondary school, vocational upper secondary school, vocational education	46.26	40.34	51.89	35.47
Short-cycle higher education	4.61	3.79	8.49	2.91
Medium-cycle higher education, bachelor	13.59	17.01	8.02	17.44
Long-cycle higher education and research education	6.70	2.98	5.66	2.33
Unknown	1.44	1.14	0.94	2.33
Sector				
Military	0.82	0.01	0.47	0.00
Leader at top level in companies, organizations and the public sector	6.36	1.63	4.72	1.16
Work that requires skills at the highest level within the area	14.89	11.03	12.74	11.05
Work that requires intermediate skills	12.78	16.94	18.87	15.12
Office Work	4.65	17.96	4.25	15.12
Sales, service and care work	3.45	17.73	3.77	18.60
Working in agriculture, horticulture, forestry, hunting and fishing, which requires skills at basic level	0.94	0.07	0.47	0.00
Handicraft work	16.85	0.82	14.62	1.16
Process and machine operator work, transport and civil engineering	10.13	2.89	9.91	1.74
Other work	10.99	10.26	9.43	12.21
Unknown	18.13	20.66	20.75	23.84

Note: Days sick from work are only registered if the spell last more than 14 days.

When we estimate our retirement models, we use different controls (total assets, income at age 58, educational level and sector of employment and different measures of health (days in hospital over the last 3 years, share out of the last 3 years sick from work[9], and reduction in 5 year survival probability due to diagnosis). The averages/shares for all these variables are listed in table 1 for the register sample and the SHARE sample of individuals. The distributions on educational groups and work sectors of employment in the two samples are pretty similar and the same is true for the average income level at age 58. Total assets are about 50% higher for men in the SHARE sample, which must be because of the small sample from SHARE consisting of singles.[10] When comparing the health measures, the picture is mixed depending on which gender we look at. Men in the registers seem to have spent more days in the hospital over the last 3 years, to have had more work absences over the last 3 years, and to have experienced a bigger reduction in the average survival probability 5 years out due to their diagnoses, than the men in SHARE. The opposite holds when we compare women in the register with women in SHARE. We therefore conclude that the SHARE sample in relation to health is biased negatively for women (women are relatively unhealthier) and positively for men (men are relatively healthier).

4.2 Health measures

In Denmark, administrative records of health care use are population-based and are available on a consistent basis for a number of years. For purposes of international comparison, we have linked SHARE-Denmark to administrative registers and use responses to self-rated health questions to calculate a principal component health index, the quintiles of the index is to be comparable with those in other countries.

4.2.1 Register-based health measures

From the registers we calculate three different health measures: a) days in hospital the last 3 years (DIH3), b) sickness absence from work the last 3 years[11] (SAW3), and c) reduction in survival probability 5 years out because of objective diagnosis (RIS5). The number of days in hospital over the last 3 years measures the total amount of days spent in hospital. Sickness absence from work measures

---

[9] Days sick from work are only registered if the spell last more than 14 days.

[10] Although we are using two waves of SHARE for Denmark we only have around 377 individuals to estimate on (1,674 panel-year observations), therefore the register analysis becomes especially important in order to identify effects.

[11] Measured as a share out of the last 3 years (not days). Only spells of work absence longer than two weeks are registered, because this is when eligibility for public sickness benefits begins and medical evidence in the form of a note from own physician is required. When aggregating over two years this censoring means that an equivalent number of days absence comprised of shorter spells compared to longer spells will be counted differently. Intermittent brief periods of absence will be omitted.

the accumulated share a person was sick from work each year over the last 3 years. If it adds up to 150, then the person was sick from work on average 50% of the time (150/3). The reduction in survival probability 5 years out is calculated as a linear probability model (LPM) with 100 different objective diagnoses[12] controlling for age, gender and year. The beta coefficients on the 100 diagnoses, which measure the reduction in survival probability 5 years out, are used to group people into health groups. All three measures of health are divided into 3 groups.

Our three administrative register-based health measures have several features in common. They are objective in the sense of requiring official registration because of associated receipt of transfers or health care. Sickness absence has a demand component to it, because it is your own choice to report in sick.[13] Not being self-assessed as a notice from a doctor is required, they are not subject to justification bias whereby individuals, in response to survey questions may end up justifying their work status based on health conditions. Such self assessments alter the estimates and the responses may be biased. However, none of our measures are of health per se.

Absence from work may have multiple causes, despite physician health certification, and these absence reasons may be related to preferences for leisure other than through health status. Hospitalization requires individuals to initiate contact with the public health care system by visiting a primary care physician, who acts as gatekeeper for hospital referral. The number of contacts with primary care physicians is registered, but we chose not to use that as a health care measure because it is obviously to some extent demand-driven care. Doctors referral filters out those for whom no further specialist investigation or treatment is needed.

Table 2 shows descriptive statistics for the three health measures from registers and their mutually correlation. The three measures span different shares of the estimation sample–DIH3 spans 19.1%, SAW3 spans 21.3% and RIS5 spans 8.2%. Because they do not span a larger share of the population, it is not possible to create deciles or quartiles. Instead we have split the share that is affected by the three respective measures into 2 equal sized groups–poor and bad (se "share all" column).

---

[12] Most dominant diagnose (at the end of each year) if more than 1 diagnose.

[13] Actually all three will have a demand component to them, because you often decide on your own, if you want to go to the doctor to be examined.

Table 2: Correlation between health measures in registers, estimation sample

	min – max value	Observations	Share all	Share men	Hospital days	Sick from work	Reduction in 5 year survival due to BETA
Health group according to days in hospital last 3 years (DIH3)							
All		1,072,499	100.0	0.57	1.3	3.2	-0.77
Good	0 – 0	867,480	80.9	0.56	0.0	1.7	-0.56
Poor	1 – 3	105,662	9.9	0.58	1.6	5.5	-1.13
Bad	4 – 418	99,357	9.3	0.59	12.8	13.6	-2.18
Health group according to sick from work last 3 years (SAW3)							
All		1,072,499	100.0	0.57	1.3	3.2	-0.77
Good	0 – 0	844,199	78.7	0.58	0.6	0.0	-0.64
Poor	1 – 7	109,904	10.2	0.53	1.5	3.8	-0.98
Bad	8 – 159	118,396	11.0	0.51	6.4	25.3	-1.46
Health group according to BETA coefficient (RIS5)							
All		1,129,366	100.0	0.57	1.4	3.2	-0.76
Good	0 – 0	1,037,070	91.8	0.56	1.2	2.9	0.01
Poor	-0,2 – -5	46,625	4.1	0.59	3.1	5.6	-2.68
Bad	-5,1 – -73,6	45,671	4.0	0.65	4.9	6.2	-16.38

Note: The table show how the three health measurements we use from the registers are correlated.

Overall the estimation-sample consists of 57% men. This share is pretty stable across the 3 health level groups of DIH3. For SAW3 the share of men is lower when going from "Good" to "Bad" indicating that more women than men have long spells of sickness from work. The opposite pattern applies to RIS5. Here men represent a relatively bigger share of the group "Bad," indicating that relatively more men receive serious diagnoses that reduce the survival probability by more than 5% points.

If we look at the correlation between the three health measures, there seems to be a strong correspondence between the groups "Good", "Poor", and "Bad" across the three measurements.

Table 1a in the Appendix A shows the Pearson correlation coefficient between the 3 health measures on a continuous scale. The strongest correlation between the three health measures is between "days in hospital" and "absence from work," where the coefficient is 0.31. The relation between "the 5 year survival probability" and the two other health measures is "only" between -0.6 and -0.11. All correlations are highly significant.

4.2.2 SHARE-based health measures

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel database of microdata on health, socio-economic status and social and family networks. This database includes more than 55,000 individuals from 20 European countries aged 50+. The sample covers non-institutionalized individuals and their spouses. There have been 4 waves so far: wave1=2004, wave2=2006/7, wave3=2008/9, and wave4=2010. In this paper we use wave 1 and 2 for Denmark, which respectively contained 916 and 1.409 observations aged 50-64 (including spouses). Since our estimation sample spans the ages 58-64/66 and our register sample "only" covers 50% of the Danish population, we are only capable of finding 384 of these individuals in our sample. SHARE wave 1 and 2 contain multiple questions on cognitive, mental, and physical health–we use the questions on physical health to create a health index.

To maintain as much comparability across countries as possible, we calculate a continuous index that divides individuals into health quintiles–we use the index that Poterba, Venti, and Wise (2011) have used in several contexts. This simple index is the 1st principal component of many health indicators. In table 2a in the Appendix A we show the physical health questions and their respective loadings on the first principal component. All questions are coded 0 or 1, and 1 indicates worse health.

Table 3 shows the correlation between the SHARE health index and the three health groups formed from the registers (shown in table 2). Overall there seems to be a high consistency between the share health index and the three health measures from the registers, so even though they capture different dimensions of health, they also seem to capture some common health indicator. The Pearson correlation coefficient in table 1a in the Appendix A also supports this finding.

Table 3: Correlation between health measures in registers and SHARE, estimation sample

Health group based on PC SHARE (PCS)	min – max value Principal component	Observations	Share all	Share men	Hospital days	Sick from work	Reduction in 5 year survival due to BETA
All		1,674	100.0	0.59	1.2	2.5	-0.81
0 – Best	-0,67 – -0,65	337	20.1	0.62	0.5	0.9	-0.67
1	-0,64 – -0,57	332	19.8	0.59	0.2	1.0	-0.21
2	-0,57 – -0,45	334	20.0	0.59	1.1	2.3	-0.83
3	0,45 – -0,24	338	20.2	0.65	2.3	3.1	-0.81
4 – Worst	-0,24 – 5,05	333	19.9	0.47	2.1	5.2	-1.53

Note: The table show how the three health measurements we use from the registers are correlated with the health measurement from SHARE.

From table 3 we also see that the score on the first principal is not very different for the first 4 groups (spans -0.67 to -0.24) while there is a big difference in the score for the group 5 (spans -0.24 to 5.05). This indicates that the health status for the top 80% is stable, while some in the bottom 20% seem to have much worse health. This is in good accordance with the health measures from the registers, where only between 10-20% are affected by the different health measures we use.

4.3 Weighting different pathways to retirement

Eligibility for some programs is more easily observed than for others. PEW eligibility can be calculated from observed UI and PEW contribution histories over the relevant periods. Benefits are a function of retirement age, cohort and private pension drawdown while OAP eligibility is a simple function of age and birth cohort, and entitlement depends upon marital status. The DI pathway is more problematic in that everyone potentially is eligible for benefits, but only a few people would consider applying for DI, and the researcher does not observe a priori who these few people are.

For estimation it is necessary to accommodate the extent to which DI is a relevant route to retirement for each individual. Assigning a probability to each person's eligibility for DI as opposed to other known eligibility routes allows DI to be incorporated into a single OV measure. In calculating DI eligibility probabilities, there are different assumptions one can plausibly make regarding which explanatory variables to use and whether to consider probabilities of award (flows) or probability of receipt (stocks).

Two DI probabilities capture the basic alternative measures and each has its merits. A stock measure calculates the population share aged 50-66 that is receiving DI in a given calendar year. This calculation includes not only age cohorts but also those who entered the program while young. An age-specific flow is the share of workers of a given age who enter DI, and this obviously reflects admission to the program but understates the extent to which DI is a realistic option for future ages.

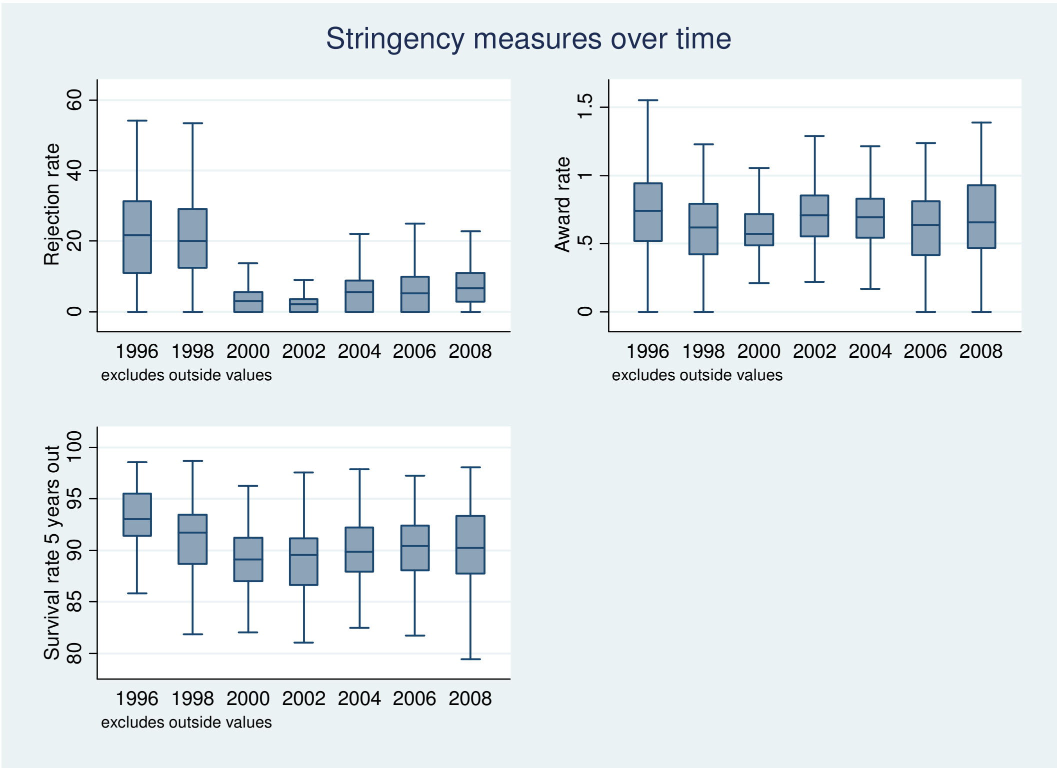
In practice we follow a combined approach. Stock probabilities are calculated ages 50-66 by gender, year and schooling for use in weighting DI in OV calculations. A downside with the stock measure is that stringency reforms during the sample period are washed out by the large number of recipients awarded while young. We therefore also estimate age-specific DI award flow probabilities in order to incorporate stringency reforms.

4.3.1 Three DI stringency measures

One way to capture the effect of the DI program on labor supply is by exploiting the variation in the awarding process between municipalities over time. We do that by defining 3 measures that are all related to the stringency of the DI program. The three measures are: the rejection rate, the population award rate and the 5-year survival rate for those awarded DI pension. The rejection rate measures the share of new applicants in each municipality[14] each year who were not awarded DI pension. The population award rate is number of awards relative to the population size in each municipality in a given year. The third stringency measure measures the survival rate 5 years out for those who were awarded DI pension in a given municipality in a given year. When the survival probability for those awarded goes up, it indicates that the stringency has gone down (more healthy people are being awarded DI), and therefore the probability of being awarded DI increases.

Figure 5 shows the development in the stringency measures from 1996 to 2008.

Figure 5: Development in stringency measures 1996-2008, percentage



Note: The figure show the development over time in the three DI stringency measures that we use. The rejection rate is the rejection rate of DI applicants in each municipality. The award rate is the number of awards in each municipality compared to the population. Survival rate 5 years out, is the survival rate for those who were awarded DI over time.

The rejection rate decreases drastically after the 1998 filtering reform, which demanded that all other possible options should be tried before a person could apply for DI pension.[15] The large decrease in the

[14] Both municipalities and special boards could process an application, but we only have data for the municipality cases which is the largest share.

[15] The municipalities both prepared an application and processed the application in order to make sure that all other possible options were exhausted.

rejection rate is therefore primarily due to the fact that a lot of those who would have been rejected before the 1998 reform no longer were allowed to apply for DI pensions. This point is supported by the fact that the population award rate fell from 1996 until 2002, indicating that stringency actually increased in the period. The survival probability 5 years out for those awarded also decreased from 1996 to 2002,which also supports the conclusion that stringency went up, because those awarded were less healthy. Thus there seems to be a common pattern: Stringency increased from 1996 to 2002 and then decreased from 2002 to 2008.

4.3.2 DI models and predicted probabilities

Table 4 reproduces the flow and stock probability models. In all three models, we control for year, gender and education level. In the second flow model, we also control for age and health indicators, which improves the pseudo R2 significantly. It only makes sense to include DI stringency measures in flow models, because the stock model includes DI recipients who were awarded DI long before the application of the DI stringency measures.

Table 4: Probability model for DI based on FLOW and STOCK estimates

	STOCK model	FLOW model 1	FLOW model 2
	margins_b/Std.	margins_b/Std.	margins_b/Std.
DI reject rate by municipality and year/applicants		-0.0289***	-0.0646***
		(0.0051)	(0.0067)
DI award rate by municipality and year/population		0.5557***	0.6797***
		(0.0121)	(0.0167)
Survival 5 years out for awarded by municipality and year		0.5840***	0.968***
		(0.1168)	(0.16158)
Year	x	x	x
Gender	x	x	x
Education	x	x	x
Age dummies			x
Duration characterized as low chance to get a job because of social problems			x
Absence from work due to sickness last 3 years			x
Days in hospital 3 years back			x
Survival probability 5 years out due to beta from survival model			x
Observations	4.171.035	2.700.128	2.466.974
Pseudo R2	0.155	0.063	0.2348

All three stringency measures are significant and with the expected signs: a) the higher the average rejection rate in a municipality, the lower the probability that the individual will be awarded a DI pension, b) the higher the number of awards in each municipality compared to the municipality population, the higher the probability of being awarded a DI pension and c) the higher the survival probability 5 years out for those awarded, the higher the probability of being awarded a DI pension.

The relatively low marginal effect of an increase in the rejection rate can be due to the filtering reform in 1998, which caused the rejection rate to go down, because people who would have been rejected before the reform no longer would apply for DI pension. This large change in the "risk-set" may be the reason why the correspondence between the rejection rate and the inflow is so relatively weak.

Table 5 shows descriptive statistics for the three stringency measures and the predicted DI probabilities from the STOCK and the two FLOW models.

Table 5: Descriptive statistics, Probability of going to DI, STOCK and FLOW

	Observations	Mean	Min	Max	p1	p5	p10	p90	p95	p99
Reject rate	2,742,477	10.00	0.000	100.0	0.00	0.00	0.00	25.72	32.04	46.66
Award rate	2,752,458	0.60	0.000	9.1	0.00	0.09	0.21	0.95	1.09	1.32
Survival 5 years out	2,708,020	90.10	42.313	99.0	78.09	82.67	84.83	94.83	95.97	97.50
Probability DI Flow1	2,700,128	1.83	0.084	14.3	0.21	0.36	0.49	3.80	5.55	7.65
Probability DI Flow2	2,466,974	1.70	0.002	88.7	0.02	0.04	0.07	4.56	7.88	19.11
Probability DI STOCK	4,171,035	25.60	0.01	59.52	0.03	0.14	2.65	50.82	55.41	59.52

The average rejection rate is 10% and spans from 0% to 100%. It spans the whole range because some municipalities had very few cases each year. The average award rate is 0.6% and the range goes from 0% to 9.1%. The survival probability is on average 90.1% and spans from 42.3% to 99.0%. The DI probability varies a lot between the three models. In the two FLOW models the average DI probability is 1.83% and 1.70% while it is 25.60% in the STOCK model. The STOCK model puts much more weight on the DI pathway than the FLOW models.

The second FLOW model (FLOW2) predicts DI probabilities in a much larger range than FLOW1–(0.084-14.3) vs. (0.002-88.7)–indicating that health is important in order to differentiate probabilities across individuals.

5. Estimation and results

In this section we present results from three different estimations. First, we present the retirement model based on the SHARE sample and the health index calculated from the health questions in SHARE wave 1 and 2. Secondly, we present the retirement models based on the register sample and the health groups calculated from the register based health indicators. Thirdly, we change DI stringency to see how the reweighting of the DI pathway influences retirement.

When we look at retirement models, we are interested in the correlation between health and retirement, which we try to capture by examining the interaction of health with the economic incentive to retire (the OV measurement). We expect the marginal effect on the OV measurement to be negative, which indicates that the bigger the difference between the present value of retiring today and the present value of retiring at the point in time that gives the highest present value, the less likely it is that people will retire today. We expect people in the worst health group to retire with a higher probability than those in the best health group, and we expect the unhealthy to react less strongly to the economic incentive on the margin. To measure how the reaction to the OV on the margin, changes with health levels, we interact health with the OV measurement and expect the marginal effect to be less negative for those in the worst health groups.

Table 6 present estimates of retirement models based on health from SHARE and table7, table 4a and table 5a present estimates of retirement models based on health from register data. Each table presents estimates from models that interact OV with different measures of health. The tables have an identical structure to facilitate comparison of different health measures. Within each table, columns show estimated coefficients for separate regressions using different sets of control variables. In all four tables the first column is the simplest reference specification and includes only OV. Column 2 includes OV and a full set of controls. Column 3 includes health groups and health groups interacted with OVs. Columns 4-7 additionally include more groups of control variables.

5.1 Retirement model with health calculated from SHARE

Table 6 presents the estimation results for the model based on the SHARE sample with the SHARE health index. The health index is calculated based on the first principal component score from multiple health questions from wave 1 and 2 of SHARE. Afterwards, we split the score into 5 equal- sized groups, where group 1 is the best health group and group 5 is the worst health group.

The marginal effect on OV is negative for all but one specification. However, the OV's are only negative and significant when no controls (X's) are added to the specification. So the average individual delays or moves forward retirement age according to financial incentives, but not significantly so. Coefficients on OV's interacted with health are everywhere insignificant. This suggests that, even individuals in better health are unresponsive to incentives. Coefficients on health quintiles per se are never significant, but point estimates are always negative.

We suspect the lack of significance is due to the small sample from SHARE, which we test later.

Table 6: Retirement model with health, based on SHARE health questions

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.
OV	-0.9703***	0.1234	-1.0283***	-0.6986	-0.8712	-0.7297	-0.3575
	(0.1246)	(1.0842)	(0.2953)	(5.3146)	(6.7730)	(6.1132)	(3.0401)
PC=2			0.1155	0.2268	0.1757	0.1792	0.1709
			(0.0763)	(1.7205)	(1.3667)	(1.4990)	(1.4213)
PC=3			-0.0486	-0.0565	-0.0917	-0.1370	-0.1584
			(0.0735)	(0.4430)	(0.7205)	(1.1485)	(1.3176)
PC=4			0.1020	0.2299*	0.1857	0.1474	0.1182
			(0.0727)	(1.7436)	(1.4434)	(1.2343)	(0.9874)
PC=5			0.0789	0.1748	0.1371	0.0830	0.065
			(0.0661)	(1.3266)	(1.0679)	(0.7023)	(0.5523)
PC=2*OV			-0.2239*	-0.3761*	-0.3433	-0.3626	-0.3581
			(0.1102)	(2.8515)	(2.6665)	(3.0277)	(2.9723)
PC=3*OV			0.1126	0.2138	0.2434	0.2781	0.2999
			(0.0793)	(1.6236)	(1.8908)	(2.3221)	(2.4881)
PC=4*OV			-0.0405	-0.0933	-0.0593	-0.0234	0.0002
			(0.0866)	(0.7212)	(0.4841)	(0.2568)	(0.1672)
PC=5*OV			0.0121	0.0546	0.0857	0.1224	0.1176
			(0.0626)	(0.4270)	(0.6740)	(1.0280)	(0.9822)
Age		x		x	x	x	x
Gender		x		x	x	x	x
Total asset		x			x	x	x
Branche		x				x	x
Education		x				x	x
Income		x					x
Observations	1,559	1,466	1,559	1,559	1,481	1,466	1,466
Pseudo R2	0.074	0.248	0.091	0.245	0.239	0.265	0.267

Note: OV is the option value measurement. PC is the principal component group based on the principal component index created from SHARE health variables. PC=1 is the best health group and PC=5 is the worst health group.

5.2 Retirement model with health calculated based on register variables

We have estimated the same 7 retirement models, as we did for the SHARE sample, for all three health variables (days in hospital last three years, sick from work over the last three years and survival rate 5 years out based on diagnoses). Again we have used the DI probability based on the STOCK sample, in order for the setup to be as close to the SHARE approach as possible. Because the three health measures from the registers only span 10-20% of the estimation sample, we cannot split the sample into 5 health groups as we did with the SHARE health index.[16] Instead we split the health variables into three groups, where group 0 is the healthy group, group 100 the less healthy group, and group 200 the least healthy group (See table 2 for details.)

Table 7, table 4a, and table 5a, respectively, show the retirement models with the three different health measures (days in hospital, survival 5 years out based on data from diagnoses, and sick days from work) from the registers with different sets of controls.

Table 7: Retirement with health, based on days in hospital from registers

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.
OV	-1.2361***	-0.5445***	-1.2671***	-1.0324***	-1.0416***	-0.8125***	-0.5834***
	(0.0057)	(0.0114)	(0.0063)	(0.0102)	(0.0104)	(0.0104)	(0.0122)
Hospital=100			0.0087***	0.0123***	0.0128***	0.0125***	0.0129***
			(0.0014)	(0.0019)	(0.0019)	(0.0019)	(0.0019)
Hospital=200			0.0099***	0.0211***	0.0209***	0.0204***	0.0208***
			(0.0013)	(0.0017)	(0.0017)	(0.0017)	(0.0017)
Hospital=100*OV			0.0009	0.0027	0.0022	0.0011	0.0013
			(0.0016)	(0.0022)	(0.0023)	(0.0023)	(0.0023)
Hospital=200*OV			0.0215***	0.0286***	0.0279***	0.0266***	0.0258***
			(0.0013)	(0.0018)	(0.0019)	(0.0019)	(0.0019)
Age		x		x	x	x	x
Gender		x		x	x	x	x
Total asset		x			x	x	x
Branche		x				x	x
Education		x				x	x
Income		x					x
Observations	1,066,697	1,035,930	1,066,697	1,066,697	1,035,930	1,035,930	1,035,930
Pseudo R2	0.093	0.227	0.094	0.217	0.216	0.227	0.229

Note: OV is the option value measurement. Hospital=0 is the best health group and Hospital=200 is the worst health group. The health groups are based on days in hospital during the last 3 years.

Overall there are a number of important patterns across the three tables:

1. The marginal effect on the OV measurement is negative and strongly significant in all models– but the marginal effect is reduced when more controls are added. This means that the higher the value of the income stream from retiring in some future year compared to retiring this year,

[16] It is important to remember that the first 4 groups of health index hardly varies on the principal score calculated from SHARE, which supports the findings from the register health measures, that only between 10-20% seems to be in worse health than the rest.

the less likely it is that retirement will be chosen. Individuals choose earlier or later retirement according to financial incentives to do so, as theory would suggest.

2. Belonging to health group 100 or 200 (poor and bad) increases the retirement probability, and being in "bad" health (group=100) increases the probability relatively more than being in "poor" health (group=200).
3. The marginal effect on the interaction term (OV*health) for those in "bad" health (group=200) is positive and strongly significant in all models, indicating that they are less responsive on the margin to economic incentives.
4. The marginal effect on the interaction term for those in "poor" health (group=100) is not significant in the models based on days in hospital (table 7) and with the beta coefficient from the survival model (table 4a)–but the effect is significant in the model that uses absence from work (table 5a), and the sign is negative (opposite of what we would expect).
5. Estimated coefficients are remarkably robust across specifications including a host of different groups of control variables.

Work absence is censored in that short periods of absence are not observed. Hospitalization is a censored measure of health in that the condition needs to be deemed severe enough by a physician gatekeeper to merit hospital admission. Both of these censoring mechanisms mean that we only observe non-zero values for 20% of the population. The remaining 80% are censored and assumed to be in good health. Therefore we cannot say more about a health gradient in retirement behavior on the basis of the register-based health measures than we have access to. Notwithstanding this, our results are strikingly consistent across the three somewhat different register-based health measures. Individuals delay or move up retirement age according to financial incentives to do so and those in better health are more responsive to these incentives.

To reconcile the lack of incentive or health effects found from table 6 with the strong and consistent findings in the tables based on the register sample in tables 7, 4a and 5a, we need to think about what has changed. Health measures are different, both in their nature and span across the whole distribution. From SHARE we calculate a health index from a principal component analysis of self-reported health as opposed to our objective administrative measures. In SHARE we also manage to span the whole distribution with the health index created from SHARE health questions, but this measure does not distinguish among the worst two deciles (group 5).

The downside of this greater breadth of coverage across the health range with SHARE is the large reduction in sample size compared with administrative data (a thousand SHARE observations compared to a million administrative observations). And as we show in table 3a in the Appendix A: replicating our analysis using register-based health measures but restricting it to the SHARE sample, also reveals **insignificant** marginal effects for health and OV and health interactions. This suggests that sample size is important, at least to obtain power enough to measure differential health effects and incentive interactions at the bottom of the health distribution.

5.3 Observed, predicted and simulated retirement probabilities based on register variables

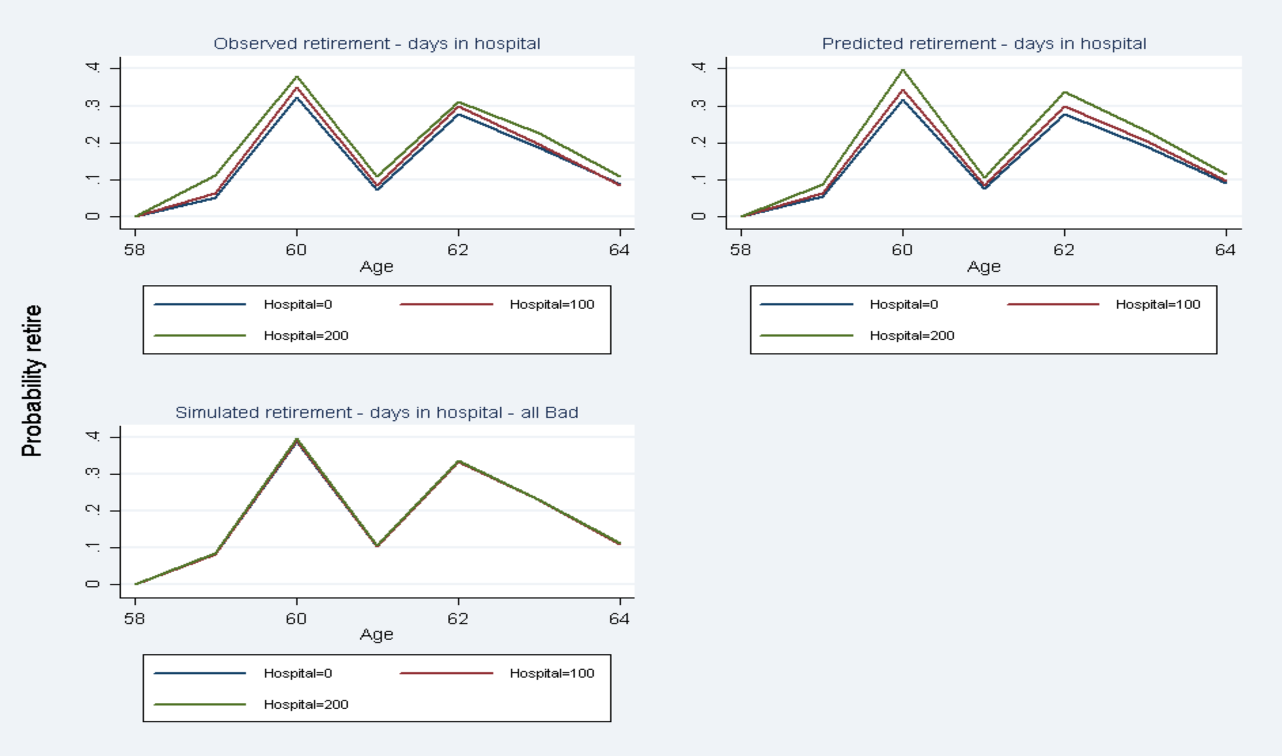
To illustrate the correlation between the size of the OV, the health level, the marginal effect on the OV, the marginal effect on the health level and the marginal effect on the interaction term (OV*Health), we predict retirement probabilities by age for the three health definitions.

First we estimated observed hazard rates split by health group for each of the three health measures, then we predicted retirement hazard rates based on the model with the full set of controls, and finally we simulated retirement as if everyone was categorized as belonging to the worst health group.

Figure 6 shows the observed, predicted and simulated hazard rates for the model with health grouped according to days in hospital. The predicted hazard rates are close to the observed hazard rates, and the ranking of the curves is also similar–the worse the health, the higher the retirement probability. If we simulate that all individuals' health is as bad as those in the worst health group (4 days or more in hospital during the last 3 years), then the curves almost collapse into one curve. This is because the three health categories (group by days in hospital) do not differ much on the control variables (not shown here).

Figure 3a in Appendix A, shows the observed, predicted and simulated hazard rates for the model with health divided according to sickness from work. Because the control variables are not as evenly distributed across this categorization of health, the curves do not collapse in the same way, when we simulate everybody to be in the worst health category (Bad). The three curves are also wider apart for both the observed and the predicted hazard rates.

Figure 6: Observed, predicted and simulated hazard rates to retirement, days in hospital



Note: The figures show the observed, predicted and simulated retirement probability for health defined by "days in hospital". Hospital=0 refers to individuals who were never in hospital, Hospital=100 refers to individuals that spent 1-3 days in hospital and Hospital=200 refers to individuals who spent more than 3 days in hospital the last three years.

The same pattern can be seen in figure 4a in Appendix A, which shows the observed, predicted and simulated hazard rates, with health divided by the beta coefficient from the survival model. Again the explanation for the collapse is the similar distribution on the controls over the three health categories (Good, Poor and Bad).

From figure 6, 3a and 4a we can conclude that those in bad health increase their retirement probability substantially compared to those in good health, and that the size of the increase is strongly dependent on which health measure we use.

5.4 Changing stringency of DI

In this section we simulate how changes in stringency measures affect the probability of being awarded DI and how that affect the retirement probability. We use the second flow model from table 4 with a full set of controls to first simulate the effect on the DI probability from changed stringency.

5.4.1 Effect on DI probability from changing the three stringency measures

Table 8 shows the simulated DI probabilities for different values of the three stringency measures. The suffix _10 and _90 refer to the 10th and the 90th percentile of the three stringency measures.

Table 8: Simulated DI probabilities for changed stringency, DI FLOW model

	Observations	Mean	Min	Max	p1	p5	p10	p90	p95	p99
Probability DI Flow2	2.466.974	1,70	0,00	88,74	0,02	0,04	0,07	4,56	7,88	19,11
Probability DI Flow2 Reject 10	2.466.974	1,79	0,00	90,05	0,02	0,04	0,07	4,87	8,30	19,87
Probability DI Flow2 Reject 90	2.466.974	1,58	0,00	88,89	0,02	0,04	0,06	4,26	7,37	18,15
Probability DI Flow2 Award 10	2.466.974	1,19	0,00	84,67	0,01	0,03	0,04	3,10	5,53	14,56
Probability DI Flow2 Award 90	2.466.974	2,04	0,01	90,18	0,03	0,07	0,10	5,52	9,25	21,60
Probability DI Flow2 Survival 10	2.466.974	1,63	0,00	88,21	0,02	0,04	0,06	4,35	7,56	18,54
Probability DI Flow2 Survival 90	2.466.974	1,75	0,00	88,94	0,02	0,04	0,07	4,71	8,11	19,56

Note: The table show the simulated changes in the probability of receiving DI if stringency were changed to the 10th and 90th percentile.

If we change the DI award/population level for all municipalities each year from the observed level (average award/population=0.6%) to the level of the 10th percentile (award/population=0.21) then the average DI probability falls from 1.7% to 1.19%. If we simulate the same relative change in the survival rate 5 years out for those awarded (from observed to the 10th percentile), then the DI probability only decreases to 1.63%, although the marginal effect in table 4 show that changing the two measures is stronger for the stringency measure "survival 5 years out". The reason for the relatively smaller change in the DI probability can be seen in table 5a in the Appendix A. From table 5a we see that the change from the observed survival rate (average survival rate=90.1%) to the 10th percentile (survival rate=84.83%) only corresponds to a 5.8% change in the survival probability, whereas the change from the observed award/population to the 10th percentile corresponds to a 65% reduction in the award/population measure (from 0.6% to 0.21%).

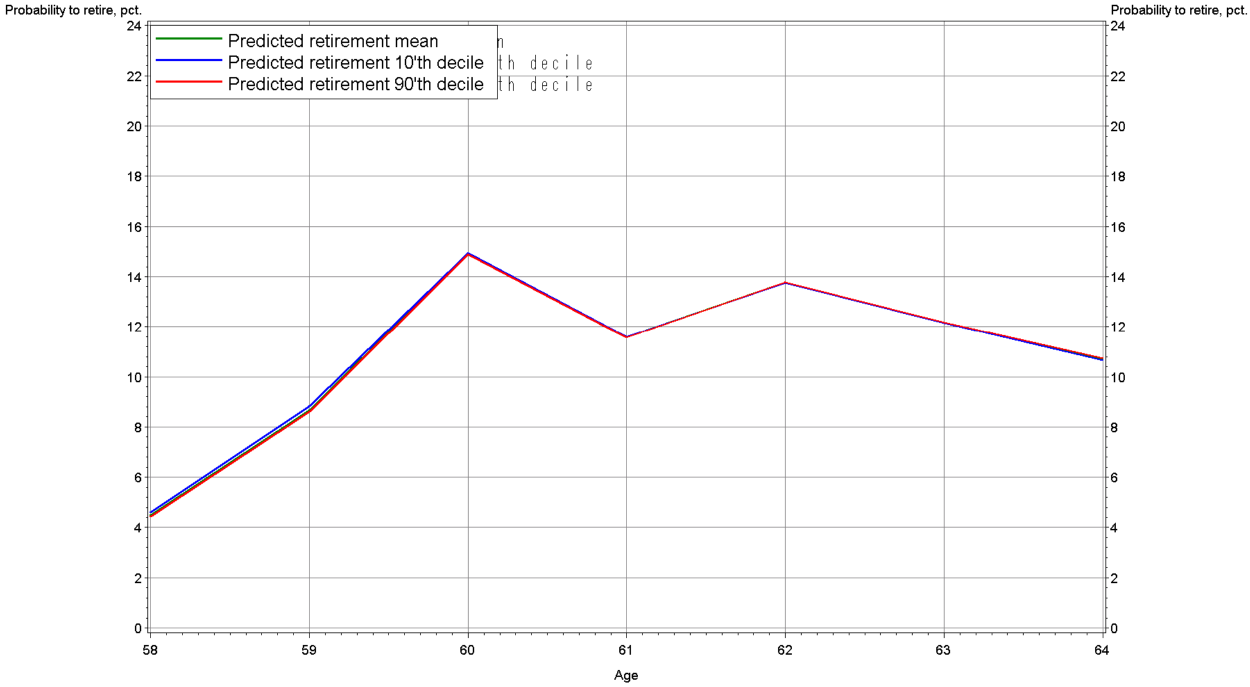
From table 4 we know that the rejection rate is negatively related to the probability of being awarded a DI pension, so therefore a similar change in stringency would be to increase the rejection rate from the observed value (average rejection rate=10%) to the 90th percentile (average rejection rate=25.67%). When we do that, the DI probability falls from 1.7% to 1.58%. This change in the DI probability is also much smaller than the change from the modified stringency, even though the rejection rate is increased by 156.7% (reduced by 39%). This is because the marginal effect on the DI probability from changing the rejection rate is much lower than the marginal effect of changing the award/population rate (see table 4).

5.4.2 Effect on retirement probability from changing (award/population) stringency

To illustrate the effect of increased stringency we simulate the effect on the retirement probability from a change to the award/population rate in each municipality-year. We change the average award/population rate from 0.6% down to 0.21% (the 10th percentile) and up to 0.95% (the 90th percentile). These two changes in (award/population) stringency respectively reduce the DI probability down to 1.19% and up to 2.04% (from the average of 1.7%).

Figure 7 shows the effect on the retirement probability from changing the (award/population) rate in model no. 1 in table 7 from the mean to the 10th and the 90th percentile.

Figure 7: Simulated retirement probabilities with changed award/population stringency, after the 1999 reform of the PEW



Note: The figure shows the effect on the retirement probability of changing the DI (award/population) rate from the mean of 0.6% to the 10th percentile (0.21%) and the 90th percentile (0.95%).

The three curves almost overlap completely, indicating that DI stringency has a very modest effect on retirement probability in the age group 58-64. One major explanation for this is the existence of the PEW program which increases retirement substantially when it becomes possible at age 60, and since about 75% of the sample is eligible for PEW at age 60, then DI becomes less important in this age group. A comparison of the observed retirement hazard rates in figure 1 (to PEW) and figure 2 (to DI) also indicates that DI has a minor role to play in retirement among the older working population.

6. Conclusions

Social security programs have long been shown to provide strong incentives for older workers to leave the labor market. Among social security programs, DI has shown to be an important route to early retirement for older workers in other countries. Its importance depends upon health conditions, labor market opportunities, stringency of awards and generosity of benefits from DI compared to other social security programs. In this paper we control for health conditions in a variety of ways in order to analyze the effect of social security incentives on the age of retirement, and we also assess the importance of the disability insurance program for retirement.

We used data for Denmark from SHARE combined with administrative registers and sampled 50% of individuals aged 58-67 during 1996-2008. This period spans major reforms to pension programs: early retirement for insured workers was made less generous, old age pension age was reduced and disability insurance awards were made more stringent. We set up an OV model of retirement age to allow for different exit routes from the labor market and characterized these changes in stringency and financial incentives.

Our results are strikingly consistent across three different health measures based on administrative register data: length of hospitalization, diagnosis severity, sickness and absences from work. Individuals postpone or bring forward retirement age according to financial incentives to do so, and those individuals in better health are more responsive to these incentives. Controlling for these measures of health, disability insurance stringency changes have modest effects within our observed range of variation.

We also estimate the retirement models based on SHARE health information in order to be able to compare our results with other countries. Results from the much smaller sample of SHARE respondents, based on a health index constructed from several self-assessed measures, are almost everywhere insignificant. Sample size for SHARE-DK probably does not provide enough data to measure differential health effects and incentive interactions. This is confirmed when we replicate these findings using administrative health measures for the SHARE sample.

Disability insurance award stringency has a small but significant role to play in early exit. The reason for this is that the early retirement program PEW offers strong incentives for early retirement. The findings on the importance of DI stringency may be specific to the Danish reform context analyzed, because

Denmark experienced larger changes to its alternative social security programs compared to other countries, where changes to DI were more modest and occurred over time. Nevertheless, disability insurance always needs to be considered in the broader context of alternative routes and has an important function as a residual program for individuals who lack eligibility to enroll in other programs. Individuals in poor health are less responsive to retirement incentives, and some of those individuals will probably apply for DI before other social security programs become available.

Appendix A: Tables and figures

Figure 1a: Retirement hazard rates to all states before and after 1999 PEW reform 1999, 1992-2009

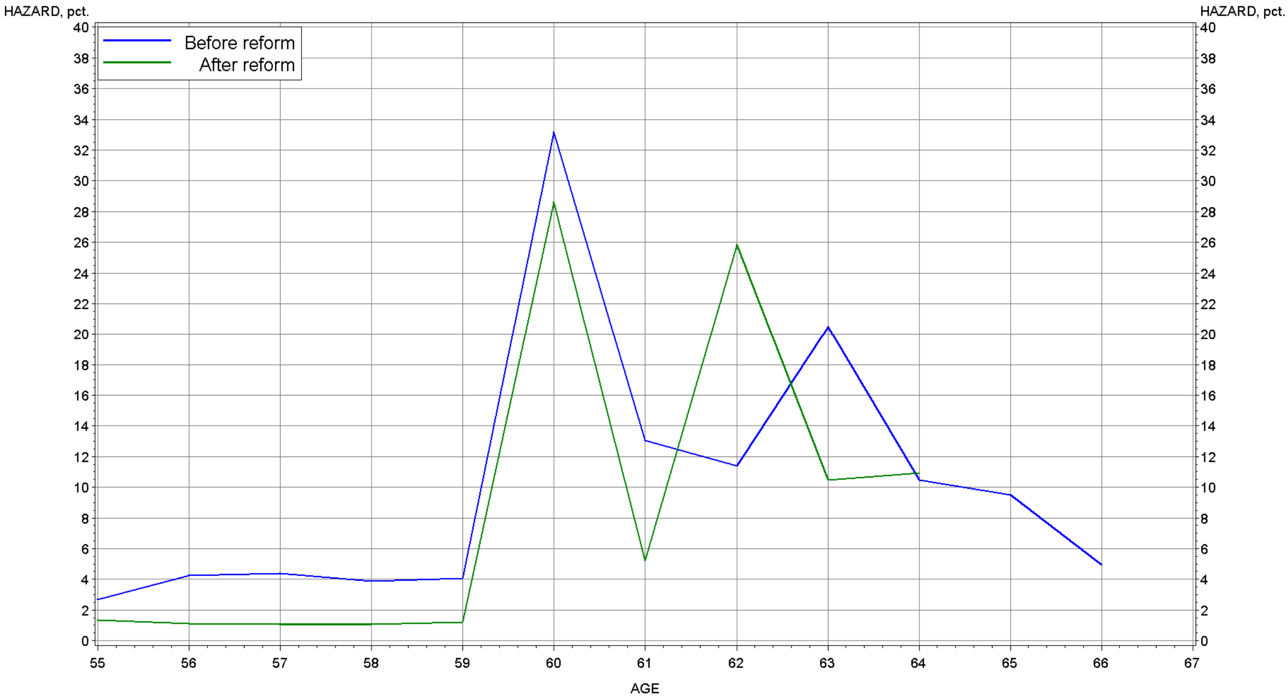


Figure 2a: Option value before and after the 1999 PEW reform by age, 1995-2008, DI

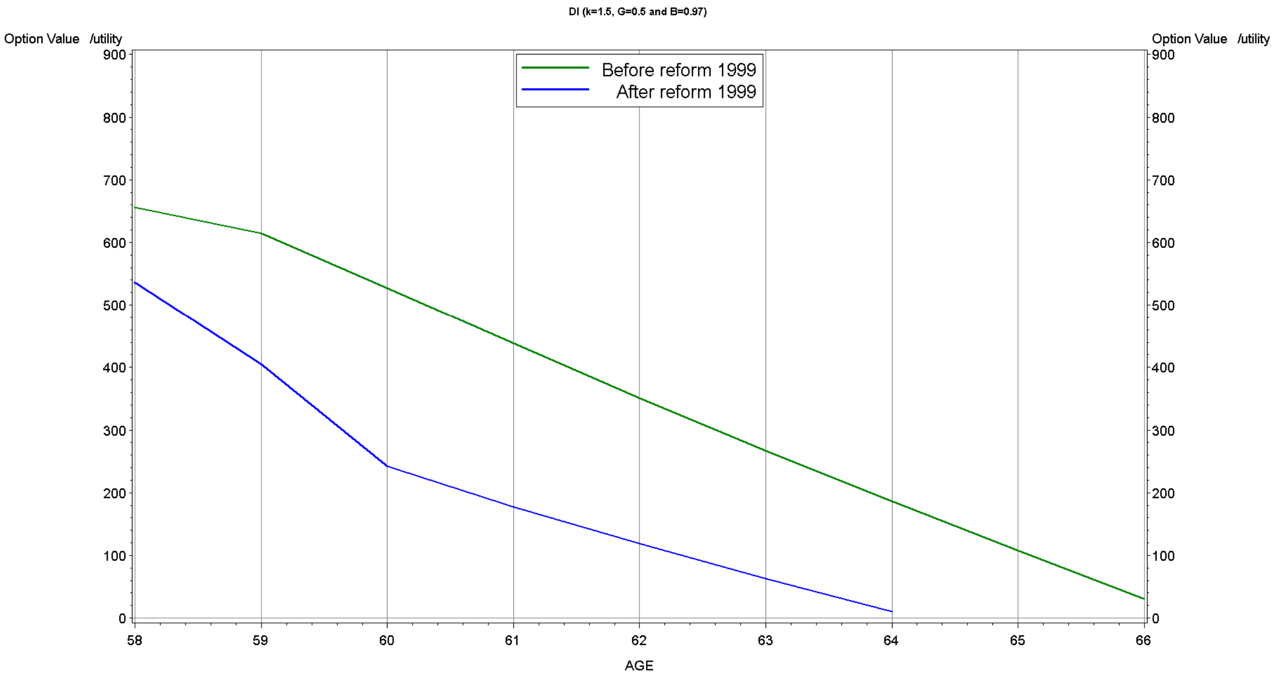


Table 1a: Correlation between health measures, register and SHARE sample

	Sickness absence last 5 years	Days in hospital last 5 years	Beta from survival model-diagnosis	SHARE health index
Observations registers				
Sickness absence last 5 years	1,130,224	1,130,224	1,129,366	
Days in hospital last 5 years	1,130,224	1,130,224	1,129,366	
Beta from survival model-diagnosis	1,129,366	1,129,366	1,129,366	
Observations SHARE				
Sickness absence last 5 years	1,674	1,674	1,672	1,674
Days in hospital last 5 years	1,674	1,674	1,672	1,674
Beta from survival model-diagnosis	1,672	1,672	1,672	1,672
SHARE health index	1,674	1,674	1,672	1,674
Pearson correlation registers				
Sickness absence last 5 years	1.000	0.309	-0.059	
Days in hospital last 5 years	0.309	1.000	-0.113	
Beta from survival model-diagnosis	-0.059	-0.113	1.000	
Pearson correlation SHARE				
Sickness absence last 5 years	1.000	0.194	-0.045	0.101
Days in hospital last 5 years	0.194	1.000	-0.059	0.084
Beta from survival model-diagnosis	-0.045	-0.059	1.000	-0.101
SHARE health index	0.101	0.084	-0.101	1.000
Probability no correlation registers				
Sickness absence last 5 years		0.000	0.000	
Days in hospital last 5 years	0.000		0.000	
Beta from survival model-diagnosis	0.000	0.000		
Probability no correlation SHARE				
Sickness absence last 5 years		0.000	0.065	0.000
Days in hospital last 5 years	0.000		0.016	0.001
Beta from survival model-diagnosis	0.065	0.016		0.000
SHARE health index	0.000	0.001	0.000	

Table 2a: First principal component and health questions from SHARE

Variable SHARE	Question	Observations	Share of population	Share men	Avg. Age	Principal component score
ALL		532	100,0	0.5	63.3	-0.14
PH004_A	LONG TERM ILLNESS YES	268	50.4	0.5	63.4	0.21
PH010D1	bothered by: pain in back, knees, hips or other joint	214	40.2	0.5	63.2	0.15
PH006D2	doctor told you had: high blood pressure or hypertension	172	32.3	0.5	63.5	-0.01
PH011D2	drugs for: high blood pressure	172	32.3	0.5	63.8	0.08
PH006D8	doctor told you had: arthritis	137	25.8	0.4	63.4	0.3
PH006D3	doctor told you had: high blood cholesterol	100	18.8	0.6	63.4	-0.05
PH048D6	difficulties: stooping, kneeling, crouching	99	18.6	0.4	63.7	0.86
PH048D6	difficulties: stooping, kneeling, crouching	99	18.6	0.4	63.7	0.86
PH006DOT	doctor told you had: other conditions	91	17.1	0.5	63.3	0.11
PH048D4	difficulties: climbing several flights of stairs	89	16.7	0.4	64	0.99
PH048D4	difficulties: climbing several flights of stairs	89	16.7	0.4	64	0.99
PH011D1	drugs for: high blood cholesterol	79	14.8	0.6	63.8	0.12
PH010D6	bothered by: sleeping problems	74	13.9	0.4	63.1	0.42
PH010D3	bothered by: breathlessness	72	13.5	0.6	63.3	0.45
PH048D3	difficulties: getting up from chair	72	13.5	0.4	63.8	0.95
PH048D3	difficulties: getting up from chair	72	13.5	0.4	63.8	0.95
PH048D9	difficulties: lifting or carrying weights over 5 kilos	68	12.8	0.3	63.7	1.22
PH048D9	difficulties: lifting or carrying weights over 5 kilos	68	12.8	0.3	63.7	1.22
PH010D5	bothered by: swollen legs	67	12.6	0.5	63.3	0.62
PH010D10	bothered by: stomach or intestine problems	56	10.5	0.4	63.6	0.6
PH005_A	SEVERLEY LIMITED ACTIVITIES 6 MONTH	55	10.3	0.5	63.2	1.22
PH011DOT	drugs for: other	52	9.8	0.4	63.2	0.07
PH048D2	difficulties: sitting two hours	51	9.6	0.4	63.2	1
PH048D2	difficulties: sitting two hours	51	9.6	0.4	63.2	1
PH011D7	drugs for: joint pain	50	9.4	0.3	63.3	0.63
PH010D9	bothered by: dizziness, faints or blackouts	47	8.8	0.6	63.1	0.67
PH011D3	drugs for: coronary diseases	46	8.6	0.7	63.4	0.5
PH011D8	drugs for: other pain	46	8.6	0.4	63.5	0.54
PH006D10	doctor told you had: cancer	42	7.9	0.5	64.5	-0.01
PH006D13	doctor told you had: cataracts	34	6.4	0.4	64.1	0.01
PH006D5	doctor told you had: diabetes or high blood sugar	34	6.4	0.6	64.3	0.02
PH049D12	difficulties: doing work around the house or garden	33	6.2	0.3	64.4	1.85
PH006D11	doctor told you had: stomach or duodenal ulcer, peptic ulcer	32	6	0.5	63.8	0.17
PH011D5	drugs for: asthma	32	6	0.4	64.1	0.4
PH006D7	doctor told you had: asthma	31	5.8	0.5	63.9	0.17
PH006D4	doctor told you had: stroke	30	5.6	0.7	64.7	0.75
PH048D8	difficulties: pulling or pushing large objects	28	5.3	0.3	64	2.16
PH048D8	difficulties: pulling or pushing large objects	28	5.3	0.3	64	2.16
PH006D6	doctor told you had: chronic lung disease	27	5.1	0.5	63.8	0.67
PH011D13	drugs for: stomach burns	27	5.1	0.5	63.4	0.16
PH011D4	drugs for: other heart diseases	26	4.9	0.6	63.5	0.72
PH011D6	drugs for: diabetes	26	4.9	0.6	64.4	-0.02
PH011D10	drugs for: anxiety or depression	25	4.7	0.2	63.1	0.71
PH048D5	difficulties: climbing one flight of stairs	25	4.7	0.4	63.9	2.24
PH048D5	difficulties: climbing one flight of stairs	25	4.7	0.4	63.9	2.24
PH010D2	bothered by: heart trouble	24	4.5	0.7	62	0.82
PH048D7	difficulties: reaching or extending arms above shoulder	24	4.5	0.4	63.5	1.78
PH048D7	difficulties: reaching or extending arms above shoulder	24	4.5	0.4	63.5	1.78
PH011D9	drugs for: sleep problems	22	4.1	0.3	63.8	0.85

Variable SHARE	Question	Observations	Share of population	Share men	Avg. Age	Principal component score
PH048D1	difficulties: walking 100 metres	22	4.1	0.4	64	2.18
PH048D1	difficulties: walking 100 metres	22	4.1	0.4	64	2.18
PH049D1	difficulties: dressing, including shoes and socks	22	4.1	0.7	64.2	2.06
PH010D11	bothered by: incontinence	21	3.9	0.3	64.5	1.04
PH010D4	bothered by: persistent cough	21	3.9	0.6	63.2	0.61
PH010D8	bothered by: fear of falling down	20	3.8	0.5	64.7	1.08
PH006D9	doctor told you had: osteoporosis	17	3.2	0.2	63.4	0.77
PH010DOT	bothered by: other symptoms	17	3.2	0.4	63.6	0.4
PH011D11	drugs for: osteoporosis, hormonal	17	3.2	0.1	63.9	0.25
PH049D8	difficulties: preparing a hot meal	17	3.2	0.4	64.6	2.26
PH048D10	difficulties: picking up a small coin from a table	15	2.8	0.4	64.2	2.54
PH048D10	difficulties: picking up a small coin from a table	15	2.8	0.4	64.2	2.54
PH049D9	difficulties: shopping for groceries	15	2.8	0.3	64.5	2.93
PH049D7	difficulties: using a map in a strange place	13	2.4	0.2	65.1	1.66
PH049D3	difficulties: bathing or showering	12	2.3	0.3	64.4	3.05
PH006D1	doctor told you had: heart attack	10	1.9	0.5	64.2	1.73
PH049D13	difficulties: managing money	10	1.9	0.5	64.2	1.73
PH010D7	bothered by: falling down	9	1.7	0.7	64.7	0.91
PH049D5	difficulties: getting in or out of bed	9	1.7	0.6	63.8	2.73
PH011D14	drugs for: chronic bronchitis	8	1.5	0.8	65.4	1.5
PH049D4	difficulties: eating, cutting up food	7	1.3	0.4	63.4	2.7
PH049D6	difficulties: using the toilet, incl getting up or down	7	1.3	0.3	63.1	2.92
PH006D14	doctor told you had: hip fracture or femoral fracture	6	1.1	0.3	63.8	0.48
PH011D12	drugs for: osteoporosis, other	6	1.1	0.2	62.7	0.27
PH049D2	difficulties: walking across a room	6	1.1	0.3	63.2	4.43
PH049D11	difficulties: taking medications	5	0.9	0.4	66	3.04
PH006D12	doctor told you had: parkinson disease	4	0.8	0.8	63.5	1.51
PH049D10	difficulties: telephone calls	4	0.8	0.3	64.8	3.05

Table 3a: Retirement with register health on SHARE sample

	Registers without controls	SHARE without controls	Registers with full controls	SHARE with full controls
	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.
OV	-0.9142***	-1.0283***	0.2551	-0.3575
	(0.1367)	(0.2953)	(2.1114)	(3.0401)
Hospital=100	0.0798*		0.1305	
	(0.0365)		(1.0652)	
Hospital=200	0.0962**		0.1890***	
	(0.0295)		(1.5410)	
Hospital=100*OV	-0.0531		-0.1212	
	(0.0443)		(0.9926)	
Hospital=200*OV	-0.0241		-0.0610	
	(0.0294)		(0.5001)	
PC=2		0.1155		0.1709
		(0.0763)		(1.4213)
PC=3		-0.0486		-0.1584
		(0.0735)		(1.3176)
PC=4		0.1020		0.1182
		(0.0727)		(0.9874)
PC=5		0.0789		0.0650
		(0.0661)		(0.5523)
PC=2*OV		-0.2239*		-0.3581
		(0.1102)		(2.9723)
PC=3*OV		0.1126		0.2999
		(0.0793)		(2.4881)
PC=4*OV		-0.0405		0.0002
		(0.0866)		(0.1672)
PC=5*OV		0.0121		0.1176
		(0.0626)		(0.9822)
Age			x	x
Gender			x	x
Total asset			x	x
Branche			x	x
Education			x	x
Income			x	x
Observations	1,559	1,559	1,466	1,466
Pseudo R2	0.091	0.091	0.266	0.267

Table 4a: Retirement with health based on BETA from diagnosis in survival model, registers

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.
OV	-1.2361***	-0.5445***	-1.2561***	-1.0132***	-1.0231***	-0.7946***	-0.5652***
	(0.0057)	(0.0114)	(0.0059)	(0.0097)	(0.0099)	(0.0099)	(0.0117)
Beta=100			0.0032***	0.0064***	0.0067***	0.0064***	0.0065***
			(0.0009)	(0.0012)	(0.0012)	(0.0012)	(0.0012)
Beta=200			0.0010	0.0075***	0.0076***	0.0076***	0.0078***
			(0.0008)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
Beta=100*OV			0.0001	-0.0002	-0.0006	-0.0009	-0.0009
			(0.0010)	(0.0014)	(0.0014)	(0.0014)	(0.0014)
Beta=200*OV			0.0134***	0.0144***	0.0140***	0.0134***	0.0131***
			(0.0008)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
Age		x		x	x	x	x
Gender		x		x	x	x	x
Total asset		x			x	x	x
Branche		x				x	x
Education		x				x	x
Income		x					x
Observations	1,066,697	1,035,930	1,066,697	1,066,697	1,035,930	1,066,697	1,035,930
Pseudo R2	0.093	0.227	0.094	0.216	0.216	0.227	0.228

Table 5a: Retirement with health based on sickness from work from registers

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.	margins_b/Std.
OV	-1.2361***	-0.5445***	-1.1397***	-0.9107***	-0.9232***	-0.7120***	-0.4759***
	(0.0057)	(0.0114)	(0.0061)	(0.0098)	(0.0099)	(0.0100)	(0.0119)
SYG=100			0.0702***	0.0666***	0.0670***	0.0613***	0.0605***
			(0.0015)	(0.0020)	(0.0020)	(0.0020)	(0.0020)
SYG=200			0.0787***	0.0840***	0.0843***	0.0768***	0.0767***
			(0.0015)	(0.0019)	(0.0020)	(0.0020)	(0.0020)
SYG=100*OV			-0.0568***	-0.0365***	-0.0359***	-0.0387***	-0.0370***
			(0.0017)	(0.0024)	(0.0025)	(0.0025)	(0.0025)
SYG=200*OV			-0.0270***	0.0054**	0.0060**	0.0055**	0.0066**
			(0.0015)	(0.0021)	(0.0021)	(0.0021)	(0.0021)
Age		x		x	x	x	x
Gender		x		x	x	x	x
Total asset		x			x	x	X
Branche		x				x	X
Education		x				x	X
Income		x					X
Observations	1,066,697	1,035,930	1,066,697	1,066,697	1,035,930	1,035,930	1,035,930
Pseudo R2	0.093	0.227	0.100	0.222	0.222	0.231	0.233

Figure 3a: Observed, predicted and simulated hazard rates to retirement, sickness from work

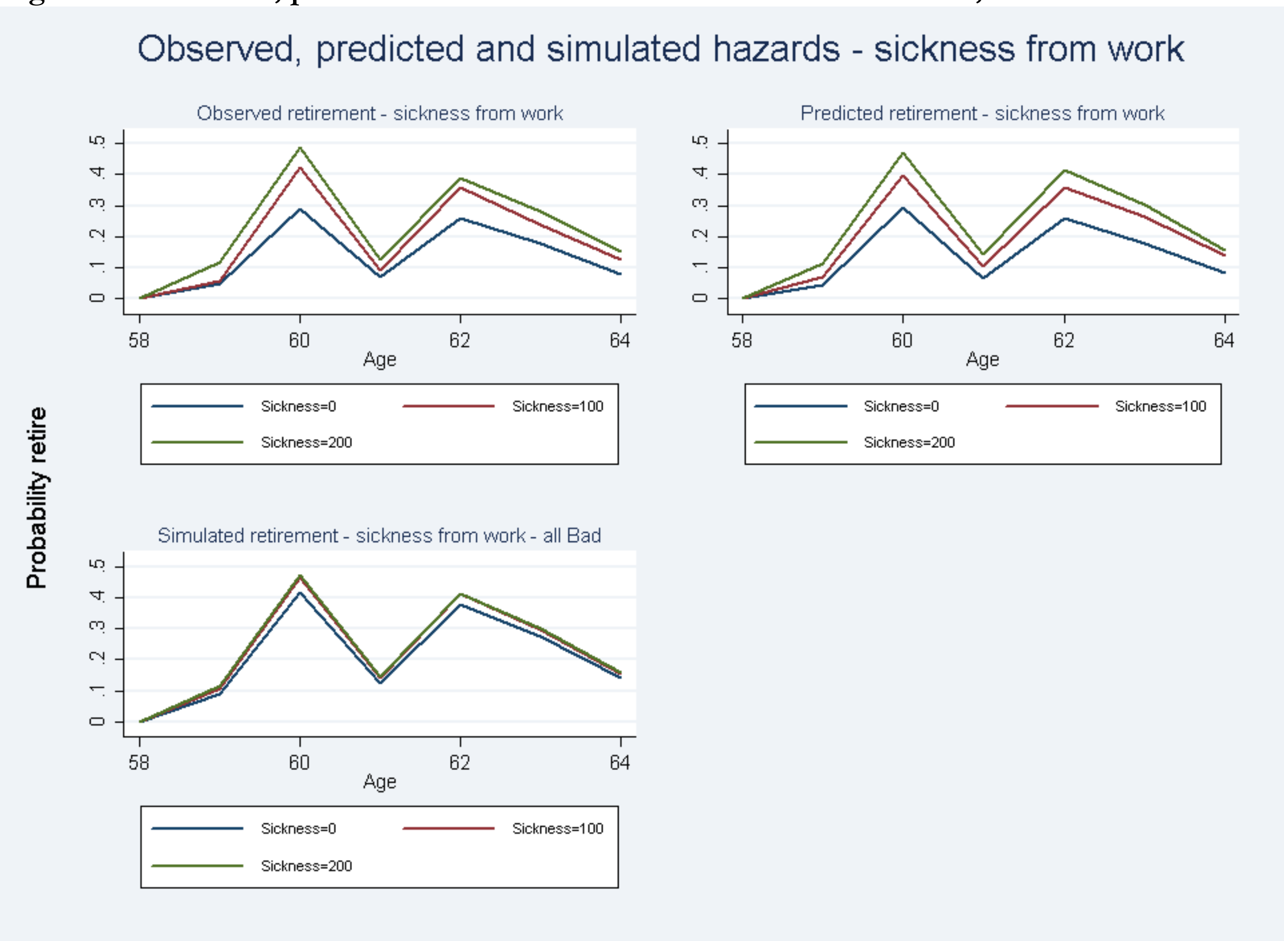


Figure 4a: Observed, predicted and simulated hazard rates to retirement, Beta from survival model

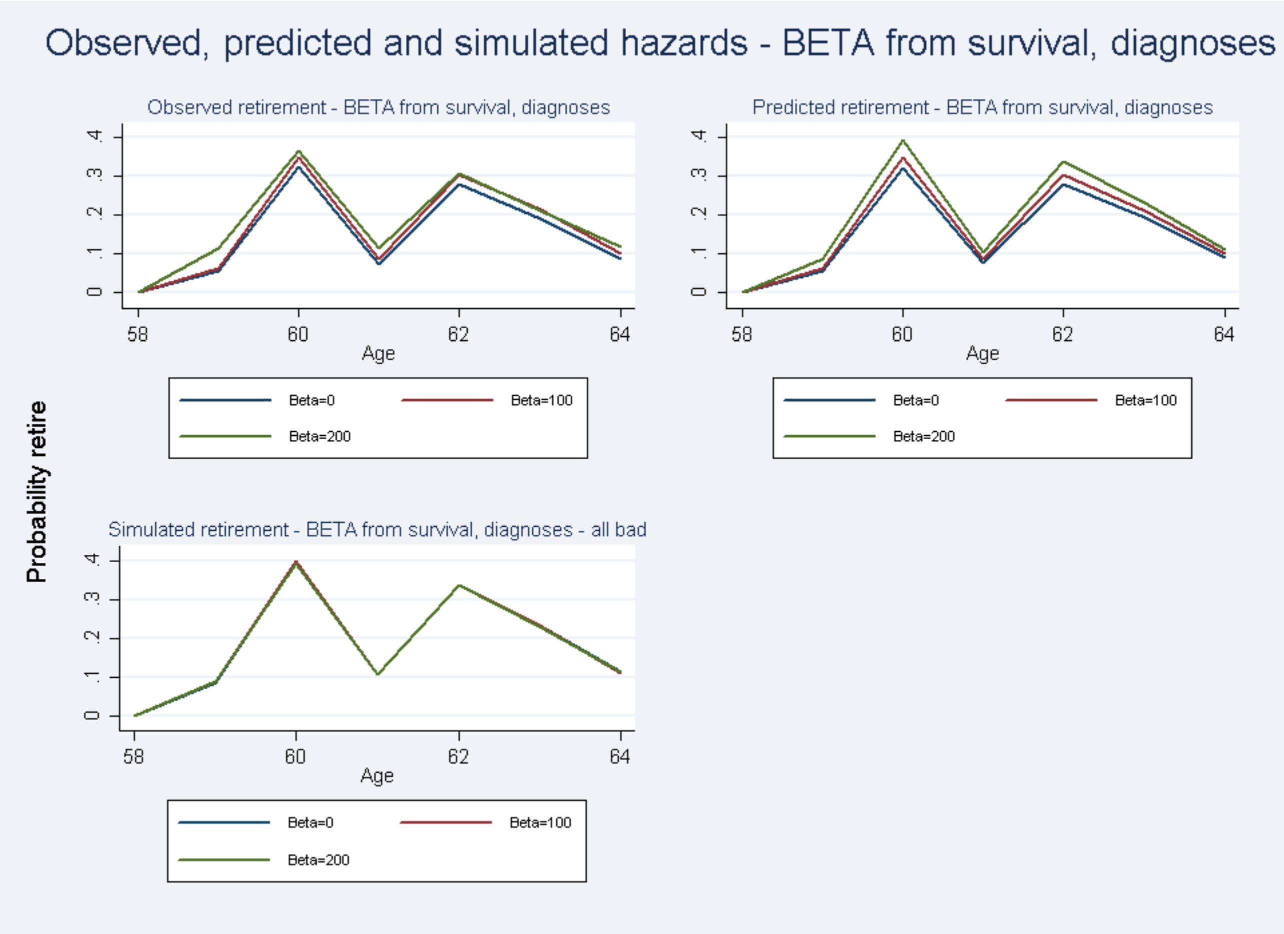


Table 5a: Retirement with health based on sickness from work from registers

	Observations	Mean	Min	Max	p1	p5	p10	p90	p95	p99
Reject rate	2.742.477	10,00	0,000	100,0	0,00	0,00	0,00	25,72	32,04	46,66
Award rate	2.752.458	0,60	0,000	9,1	0,00	0,09	0,21	0,95	1,09	1,32
Survival 5 years out	2.708.020	90,10	42,313	99,0	78,09	82,67	84,83	94,83	95,97	97,50
REJECT_10	2.752.926	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
REJECT_25	2.752.926	1,38	1,38	1,38	1,38	1,38	1,38	1,38	1,38	1,38
REJECT_75	2.752.926	14,89	14,89	14,89	14,89	14,89	14,89	14,89	14,89	14,89
REJECT_90	2.752.926	25,67	25,67	25,67	25,67	25,67	25,67	25,67	25,67	25,67
AWARD_10	2.752.926	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21
AWARD_25	2.752.926	0,42	0,42	0,42	0,42	0,42	0,42	0,42	0,42	0,42
AWARD_75	2.752.926	0,78	0,78	0,78	0,78	0,78	0,78	0,78	0,78	0,78
AWARD_90	2.752.926	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95
OVERLEVE_10	2.752.926	84,83	84,83	84,83	84,83	84,83	84,83	84,83	84,83	84,83
OVERLEVE_25	2.752.926	87,86	87,86	87,86	87,86	87,86	87,86	87,86	87,86	87,86
OVERLEVE_75	2.752.926	93,02	93,02	93,02	93,02	93,02	93,02	93,02	93,02	93,02
OVERLEVE_90	2.752.926	94,83	94,83	94,83	94,83	94,83	94,83	94,83	94,83	94,83

Note: The suffix represents the percentile at which the variable is calculated.

Appendix B: The Option Value model

1b The Option Value model

From the vantage point of each age *a* while in work, there are many possible pathways *(pa=1,..., PA)* to retirement, each with an associated utility stream *V* dependent upon age of retirement time *r*. A pathway constitutes a number of years of continued work, denoted in the first summation of equation (1), followed by the number of years receiving pension benefits specific to that pathway until death at age *A*, denoted by the second summation of equation (1). Expected utility at each future age *s* from the vantage point of each age E_a is weighted by the probability of survival to that age $p_{s|a}$ and discounted β^{s-t} back to present. While working, wage income *ω(s)* is received at each age, while retired benefit income $B_{rk}(s)$ is received at each age dependent upon pathway and age of retirement. The utility function includes a parameter for leisure *ϰ*, which scales retirement benefits relative to earnings. Both incomes in work and retirement are raised to the power γ representing risk aversion (marginal value of money):

(1) $$E_a\,\{V_{ka}(r)\} \;=\; \sum_{s=a}^{r-1} p_{s|a}\,\beta^{s-a}\big(\omega(s)\big)^{\gamma} + \sum_{s=r}^{A} p_{s|a}\,\beta^{s-a}\big(\kappa B_{rk}(s)\big)^{\gamma}$$

For each retirement pathway *pa*, the future age of retirement at which the expected discounted utility stream is maximized is denoted r^*. The comparison is between expected utility streams associated with all retirement ages until maximum age of retirement *R*. The OV of staying in work at the present age *a* compared to following eventual retirement pathway *pa* is defined as the difference between the maximum of expected utilities from future retirement ages along that pathway compared to retiring now:

(2) $$OV_{ka} \;\equiv\; \max_{a<r^*\leq R}[E_a\{V_{ka}(r^*)\}] \;-\; E_a\{V_{ka}(a)\}$$

Having defined the OV of staying in work from the vantage point of each age *a* for each retirement pathway *pa,* it remains to weight each pathway with the probability P_k so that it represents a set of relevant alternatives for each individual. An inclusive OV measure combines routes weighted by the probabilities that they are relevant as follows:

(3) $$OV_a = \sum_{k=1}^{K} P_k OV_{ka}$$

The above inclusive OV measure makes explicit the extension to the Stock and Wise (1990a, 1990b) OV approach that allow us to incorporate several different routes to retirement. This can be set in a regression framework further allowing for health status in an ad-hoc way. Consider retirement status *R* for person *i* of age *a* in health quantile *j*. This is assumed to be a function θj of exogenous individual characteristics X_{ia} and a function δ_j of inclusive OV OV_{ia}. H_j is a measure of health and ε_{iaj} is an error term:

(4) $$R_{iaj} = \sum_{j=1}^{J} \left[\theta_j X_{ia} + \delta_j OV_{ia} H_j\right] + \varepsilon_{iaj}$$

The above equation is estimated as a probit model for annual retirement. It is optimal stopping problem in that an individual remains out of the labor force once retired. Benefit collection and retirement are assumed to be synonymous. Pathways from the labor force to OAP could be direct or via DI, PEW, or a private pension drawdown. Individuals are selected ages 58-66 and must be working in first year of observation. We assume a maximum age of retirement *R* at 67 and force those who are still working at age 66 to retire at 67 on OAP. We use population life tables for survival probability *s* from age *a* published in 2009 by age and gender for ages 59-99 and impose zero survival at age *T*=100. After a year of retirement, an individual leaves the dataset. Exits from the dataset due to death, migration or change of marital status are treated as missing at random. Observations of those persons are used in estimation until the year before exit and the last observation is classified as working. Potential earnings profiles are assumed to be flat from age 58 with 1% real growth. Option value calculations assume knowledge of the pension and tax system as in place at the vantage point of observation. Individuals form expectations on the basis of that system and any future changes which had already been announced. Preference parameters β discount rate, $\varkappa$ utility of leisure and γ risk aversion are estimated freely for our most parsimonious specifications using US estimates as starting values. The grid search for parameters k and γ for β fixed at 0.97, gives k=1.5 and γ=0.5.

References

Banks J, Emmerson C, and Tetlow G (2007): *Healthy Retirement or Unhealthy Inactivity: How Important are Financial Incentives in Explaining Retirement?* Mimeo: Institute for Fiscal Studies.

Belloni M (2008): "The Option Value model in the Retirement Literature: The Trade-Off between Computational Complexity and Predictive Validity". *ENEPRI, Empirical Research Report No. 50*.

Bingley P, Gupta ND, and Pedersen PJ (2004): "The Impact of Incentives on Retirement in Denmark". In Gruber J and Wise DA (Eds.): *Social Security Programs and Retirement around the world. Micro-Estimation*. NBER. University of Chicago Press, Chapter 3.

Blau DM and Gilleskie DB (2008): "The Role of Retiree Health Insurance in the Employment Behavior of Older Men". *International Economic Review*, Vol. 49, Issue 2, pp. 475-514.

Burkhauser RV, Butler JS, and Gumus G (2003): *Option Value and Dynamic Programming Model Estimates of Social Security Disability Insurance Application Timing*. IZA Discussion Paper No. 941, November 2003.

Erdogan-Ciftci E, Doorslaer E van, and López-Nicolás Á (2011): *Does Declining Health Affect the Responsiveness of Retirement Decisions to Financial Incentives?* NETSPAR Discussion Paper 01/2011-005.

Gruber J and Wise D (2004): *Social Security Programs and Retirement around the World: Micro Estimation.* Chicago: University of Chicago Press.

Hurd M (1990): "Research on the Elderly: Economic Status, Retirement and Consumption and Saving" *Journal of Economic Literature*, Vol. 28, pp. 565-637.

Lindeboom M (2005): *Health and Work of Older Workers*. Mimeo. Free University of Amsterdam.

Lindeboom M and Kerkhofs M (2008): *Health and Work of the Elderly*. NETSPAR Discussion Paper 2007-003 (revised).

Lumsdaine R and Mitchell OS (1999): "New Developments in the Economic Analysis of Retirement". In Ashenfelter O and Card D (Eds.): *Handbook of Labor Economics*, Vol. 3. Amsterdam: Elsevier.

Lumsdaine R., Stock J, and Wise D (1991): "Three Models of Retirement: Computational Complexity versus Predictive Validity". In Wise D (Ed.): *Topics in the Economics of Aging*. Chicago: University of Chicago Press, pp. 619-660.

Manski C (1988): *Path Utility Analysis of Dynamic Choice*. Mimeo: University of Wisconsin.

Milligan KS and Wise DA (2011): *Social Security and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms–Introduction and Summary*. NBER Working Paper.

Poterba JM, Venti SF, and Wise DA (2011): *The Asset Cost of Poor Health*. HKS Faculty Research Working Paper Series RWP11-005, John F. Kennedy School of Government, Harvard University.

Rust J (1989): "A Dynamic Programming Model of Retirement Behavior". In Wise D (Ed.): *The Economics of Aging*. Cambridge: National Bureau of Economic Research, pp. 359-398.

Rust J and Phelan C (1997): "How Social Security and Medicare Affect Retirement in a World with incomplete Markets", *Economerica,* Vol. 65, No. 4, pp. 781-831.

Stern S (1999): *Approximate Dynamic Program Solutions*. Mimeo: University of Virginia.

Stock J and Wise D (1990a): "Pensions and the Option Value of Work and Retirement". *Econometrica*, Vol. 58, No. 5, pp. 1151-1180.

Stock J and Wise D (1990b): The Pension Inducement to Retire: An Option Value Analysis". In: Wise D (Ed.): *Issues in the Economics of Aging*. Chicago: University of Chicago Press.

How Does a Health Shock to Self or Partner Affect Economic Incentives to Retire?

Michael Jørgensen

Abstract

This paper exploits rich administrative data to analyze how a health shock to one-self or one's partner affects the marginal economic incentive to retire. One important challenge when trying to get unbiased estimates of the effect of health on retirement is the endogenous character of the relationship between health and retirement. To circumvent this endogeneity, the paper exploits a 1999 reform of an early retirement program and the arguably unanticipated timing of a health shock. The paper contributes to the existing literature on retirement and health in several ways. First it models the interaction between a health shock and economic incentives to retire, thereby allowing health to affect the response to economic incentives. Second, the paper exploits the unanticipated timing of a health shock, not only levels of health, which helps circumvent the endogeneity problems that health measured in levels gives rise to. Third, it uses a reform of an early retirement program that exogenously altered economic incentives without altering preferences or health in the short run. Fourth, it uses objective diagnoses from a large register sample, which allows us to analyze the effect of health shocks instead of health levels, which, due to small sample properties, are often not possible to analyze from survey samples. The main conclusions are that economic incentives have strong effects on retirement for both men and women, but that those who receive a health shock are much less responsive to economic incentives, regardless of whether the health shock strikes them or their partners. A comparison of men's and women's responses finds no differences in men and women's marginal responses to economic incentives due to a health shock to their partners–only when the health shock hits them self, then women react much stronger than men in their reduction to the economic incentive. These final results indicate that complementarity in leisure dominates substitutability when a health shock hits a partner.

1. Introduction

Since the early 2000s, the number of people receiving pensions has increased, while the number of tax payers needed to finance these pensions has decreased. This demographic change is putting pressure on Western countries in general and on the modern welfare state in particular. Both the increasing tax burden and the decreasing labor supply are forcing politicians to try to find a way to motivate the older working population and the less healthy to either retire at a later age or work more. One way of motivating people to retire later is by increasing economic incentives to keep working. However, whether that strategy will work for the less healthy population remains unknown. Knowing whether this strategy will work for the less healthy is essential for policy makers, as the demographic trend (relatively more people aged 60+) is projected to increase over the coming three to four decades.

While the traditional health of Europeans has improved, declining psychological well-being and institutional changes may explain why health factors are still a major determinant of retirement. A number of studies (see e.g. Dwyer and Hu, 2000; Samwick and Wise, 2003; Banks et al., 2007) have shown that economic incentive and health (often measured in levels) are important factors affecting an individual's decision to retire, yet fewer (see e.g. Bound and Burkhauser, 1999; Coile, 2004a; Disney, 2003) have estimated how a health shock to one-self or to one's partner influences the effectiveness of economic incentives on retirement choice. Often studies have looked at the partial effects of economic incentives and health or have used information only on individuals and not on couples, even though other studies (see e.g. Gustman and Steinmeier, 2000; Coile, 2004b; Gustman and Steinmeier, 2004; Bingley and Lanot, 2007; Casanova, 2010) have shown that the retirement decision often is a joint household decision.

This paper uses data from Denmark to examine how a health shock to the self and the partner influences the response to economic incentives (of the couple) to retire. I define a health shock as a major health change and, examine different definitions in order to test the robustness of the health shock. The empirical literature suggests that poor health reduces the capacity to work and affects labor force participation substantially (Deschryvere, 2005). It is clear from the data that individuals with poor health tend to respond less to economic incentives to delay retirement. However, the effect of a partner's poor health on an individual's decision to respond to economic incentives is more ambiguous. From a care or joint leisure perspective, I would expect a health shock to the partner to lower the labor supply (i.e., increase retirement), but from an economic or budget constraint perspective, I would expect the labor supply to increase (i.e., an Added Worker Effect (AWE)) to compensate for the decline in the partner's income.

One important challenge when trying to get unbiased estimates of the effect of health on retirement is the endogenous character of the relationship between health and retirement. Generally, unobserved preferences for important factors such as leisure or a healthy lifestyle are correlated with retirement and health (Deschryvere, 2005). To circumvent this endogeneity, I exploit a 1999 reform of the early retirement program “Post Employment Wage” (*efterløn*, PEW) and the timing of an unanticipated health shock. The PEW reform is helpful because it exogenously alters economic incentives without altering preferences or health in the short run. In addition, given that the timing of a health shock is always unanticipated, it alters the health, without altering preferences or economic incentives in the short run.

This paper contributes to the existing literature on retirement and health in four ways. First it models the interaction between a health shock and economic incentives to retire, not only the partial effects. Second, as I assume health level to be strongly correlated with unobserved preferences for leisure and a healthy lifestyle, therefore the paper exploits the timing of an unanticipated health shock, not only levels of health. Third, it uses a reform of the PEW retirement program that changed economic incentives to retire for people aged 60 and 61, ceteris paribus. Fourth, the paper uses objective diagnoses from register data, which avoid the small sample problems that many surveys face when they look at health shocks instead of health levels. The objective diagnoses also avoid the justification biases that often arise in self-reported health.

The main conclusions from this paper are:

a) Economic incentives have strong effects on retirement.
b) A health shock to one-self or to one’s partner reduces the marginal response to the economic incentive significantly.
c) The definition of the health shock is crucial for the size of the marginal effect of the interaction term between economic incentives and the health shock.
d) Men and women reduce the marginal response to economic incentives by about the same amount if their partner receives a health shock, while women react much stronger than men in their reduction to the economic incentive, when the health shock hit them self.

e) The predicted retirement probabilities show that those in the two lowest quartiles of the expected life annuity payout changed their retirement probabilities relatively less than those in the two largest quartiles, from the period before the 1999 PEW reform to after the reform.

Section 1.1 provides a literature overview, and section 1.2 briefly describes the main theories on health and labor supply (retirement). Section 2 describes the important pathways to retirement in Denmark. Section 3 presents descriptive statistics and defines the different health shocks that I use. Section 4 explains the estimation equation and identification strategy. Section 5 presents results and robustness tests. Section 6 concludes.

1.1 Literature on health shocks

This paper focuses on how a health shock affects couple's economic incentives to retire. There exists a vast literature on the subject of *health and labor supply* and *an economic incentive and labor supply* (see, e.g. Barton et al., 1980; Lazear, 1986; Stock and Wise, 1990a and 1990b; Bound and Burkhauser, 1999; Gruber and Wise, 2004; Gustman and Steinmeier, 2005). However only a few papers (see e.g. McClellan, 1998; McGarry, 2004; Coile, 2004a; Banks et al., 2007; McGeary, 2009) combine two of the following dimensions: health shock to oneself or one's partner, a couple's joint economic incentive to retire and the interaction of economic incentives and health. To my knowledge, no studies have yet combined all three dimensions–especially the interaction between economic incentives and a health shock (to self or to partner), have not been studied a great deal. As these interactions will tell us whether politicians can influence a less healthy individual's retirement decision through a change in economic incentives, this study has important implications for policymakers.

Beyond the different focus just described, there exist two main methodological approaches to understanding the relationship between health, economic incentive, and labor supply–the reduced form approach and the structural form approach. I use the reduced form approach in this paper. The reduced form literature on economic incentive and retirement uses different measurements of economic incentives. Some papers simply use lag of income or lag of wealth (see, e.g. Hagan et al., 2009; Christensen and Lamb, 2010) while others (see, e.g. Bingley et al., 2004; Coile, 2004a; Coile, 2004b; Gruber and Wise, 2004; Banks et al., 2007; Belloni and Alessie, 2009; Belloni and Alessie, 2012) have used the more sophisticated option value (OV) model developed by Stock and Wise (1990a. The OV model incorporates important economic features such as: future gains and losses, updating of information each period, survival probabilities, discount rates, disutility of work, and risk aversion.

Most studies (see Deschryvere, 2005) on health and retirement have used levels of health while only a few (see e.g. McClellan, 1998; Bound and Burkhauser, 1999; Disney et al., 2003) have used health changes (health shocks). The distinction between health measured in levels and an the unanticipated timing of a health shock is a crucial distinction in literature applying the reduced form approach, because the unanticipated timing of the health shock helps to identify a causal relationship in the reduced form settings, since health is strongly correlated with past investments in health, which correlate with unobserved factors such as genes and preferences for an unhealthy lifestyle. As couples share the same lifestyle, their health capital is strongly correlated, making it difficult to identify the effect of the partner's health level on economic incentives to retire. Using the unanticipated timing of a health shock, helps avoid this problem.

According to McClellan (1998), different types of health events have very different consequences for health insurance coverage and labor supply. "Major health events have particularly large effects on retirement decisions, and these effects go well beyond the consequences of the events for functional status". Dwyer and Hu (2000) find similar results. These findings underline the importance of the definition of health.

Many papers deal with the relationship between health and labor supply in welfare systems where health coverage is insurance-based and therefore related to being employed. In such countries, the effect of an acute health shock on labor force participation is stronger than in Denmark. Butrica et al. (2009) shows that in the US medical conditions reduce non-health spending for low-income households ages 51 to 64, suggesting that holes in the health safety net before Medicare eligibility age force some low-income people to lower their living standards or to not reduce their labor supply, so as to cover medical expenses. Datta Gupta et al. (2012) find that the effect of an acute health shock on labor force participation is stronger in the US than in Denmark. Based on these findings I do not expect the direct effect of health care spending to have a significant effect on the retirement decision in Denmark, with its universal tax financed health care system.

A number of studies (see, e.g. Samwick and Wise, 2003; Coile, 2004a; Banks et al., 2007; Börsch-Supan et al., 2005; Hagan et al., 2009; Casanova, 2010) are based on self-assessed health from surveys that have been criticized for their justification biases due to self-reported health and small sample problems. For example those who are inactive may have an incentive to report worse than actual health to justify their inactivity. Such justification biases raises potential problems of both validity and reverse causality (see, e.g. Bound, 1991; Bound and Burkhauser, 1999; Lindeboom and Van Doorslaer, 2004; Christensen and Lamb, 2010). In addition small samples puts restrictions on the possibility of examining the effect of health shocks instead of health levels, because of the much lower frequency of health shocks. However, as health levels are assumed to be related to unobservable characteristics, using them will cause biases in the estimates of the health effect on economic incentives. In this analysis I define health shocks based on objective medical diagnoses from medical records, thereby eliminating both justification biases and small sample problems. By using register data, I also avoid biases from sample selection, which is frequent in survey data, (see, e.g. Groves and Couper, 1998).

More studies (see, e.g. Gustman and Steinmeier, 2000; Coile, 2004b; Gustman and Steinmeier, 2004; Bingley and Lanot, 2007; Casanova, 2010) have shown that couples have a preference for complementarity in leisure and that they base their retirement decision on economic incentives of the household, not on economic incentives of the individual. Coile (2004b) finds that men are very responsive to their wives' financial incentives, but that women are not responsive to their husbands' incentives, and present evidence to suggest that this difference may be due to asymmetric complementarities of leisure. The finding that men respond more strongly to the retirement status of their wife's than vice versa is also supported by Gustman and Steinmeier (2000, 2004) and . As our main interest is how a health shock to self or to one's partner affects the response to economic incentives to retire, it is important that the model properly control for economic incentives of the household, not only economic incentives of the individual.

Four papers that either use health shocks instead of health levels or that explore a couple's economic incentive instead of an individual incentive are Coile (2004a), Gustman and Steinmeier (2004), Banks et al. (2007) and McGeary (2009). Coile (2004a) finds that a worker's average response to the health shocks of his or her spouse is a small added-worker effect (AWE) for men, with no significant labor supply increase for women. She finds instead that women decrease their labor supply when their husband's shock is accompanied by a loss of functioning or a reduction in life expectancy. Gustman and Steinmeier (2004) find that for a wife, a husband's retirement status has an influence on her retirement decision only if spending time in retirement with her husband is important to her. They find that, for husbands, the effect of a wife's retirement on a husband's decision to retire doubles if he likes spending time in retirement with his wife, but that there is some effect even if he does not. This is consistent with the earlier findings that the husband is more influenced by having a retired spouse than the wife is. Banks et al. (2007) finds evidence that the relationship between pension accrual incentives and retirement is different for those in good and bad health. They show that greater future pension accrual is only statistically significantly associated with a reduced likelihood of retirement for men aged 50 to 59 who are in good health. The co-efficient of this measure of financial incentive is smaller and statistically insignificant for men in poor health.[17] McGeary (2009) finds that both males and females labor supply is influenced by a health shock to self and their spouses. The four papers seem to suggest that the income-effect is dominant for men, while the care-effect is dominant for women.

Even through the research in these four papers appears to overlap this paper in some respects, none of them cover all the dimensions that considered by this paper: the joint economic retirement incentive of couples, a health shock to the self or the partner, health shocks (not health levels), objective diagnoses (not self-reported health), and the interaction between a health shock and the joint economic incentive to retire (the interaction term measures the reduction to economic incentives due to a health shock). The interaction term is especially important for policy targeting, since it will tell us if we can motivate people in less good health to supply labor. The spouse's health condition is also important since 75% of people aged 60 are married or in a similar relationship.

1.2 The relationship between health and labor supply (retirement)

In theory, we can expect two main effects on labor supply from an unanticipated health shock–we can expect a positive income effect and a negative joint leisure effect. These two effects will be different depending on whether the health shock happens to oneself or one's partner and depending on whether leisure is considered a normal good or a complementary good.[18]

In the literature on health shocks there seems to be agreement on the finding that a health shock to oneself on average decreases one's own labor supply (see e.g. Coile, 2004a; Deschryvere, 2005; Banks et al., 2007). This observation stems from the fact that a health shock reduces one's work capacity, which increases the marginal disutility from work, and it may entitle the individual to non-wage income such as disability benefits. If a person is budget constrained because he/she does not have any funds to dis-save from, then he/she might be forced to continue working to avoid a large reduction in income, and therefore he/she might not reduce his/her labor supply because of a health shock. A factor that might

[17] They do not included onset of health problems as an explanatory variable for two reasons. First, because they it is difficult for them to identify which happened first, the change in health or the exit from work. Secondly, they observe only a two year period, so the number of people experiencing the onset of a new health condition is small.

[18] Normal goods are goods for which demand increases when income increases and falls when income decreases but price remains constant. Complementary goods are goods with negative cross elasticities of demand, in contrast to a substitute good.

amplify the budget-constraint effect (the dampening of the reduction in labor supply) is if health care coverage is highly dependent on working. In that case, the individual might be forced to supply labor (job-lock effect) in order to receive/pay for health care. In Denmark, that effect is not expected to be very strong because of its universal tax-financed health care coverage.[19]

When it comes to the effect of a health shock to one's partner on one's own labor supply, the results in the literature are more ambiguous (Deschryvere, 2005), because a health shock to one's partner will create both a demand for economic compensation and a demand for care/shared leisure. The demand for economic compensation comes from the fact that the partner who suffers the health shock is expected to lower his/her labor supply, which results in lower life-time earnings. To compensate for this reduction in income, the healthy partner can choose to increase his/her labor supply (an AWE). On the other hand, the healthy partner might have a preference for joint leisure– also known as complementarity in leisure–and therefore prefers to retire together with the unhealthy spouse.[20] Those few articles that have tried to answer the effect of a health shock to one's partner on one's own labor supply by looking at health shock and not just health levels have reached different conclusions.

Because we can not measure unobservable factors such as preferences for a healthy life style, I try to identify if the health shock effect is dominated by an added-worker effect (AWE) or a care/leisure effect, by using a reform of an early retirement program that changed economic incentives to retire and by utilizing the fact that the timing of a health shocks is unknown to the individual.

[19]Datta Gupta et al. (2012) show that the average out-of-pocket medical expenditures for people aged 55-64 was $ 640. In comparison the expenditures for the same age group was only $ 152 in Denmark

[20]The complementarity in leisure can be strengthened if the person who suffers the health shock needs more care because of the health shock.

2. Retirement in Denmark

In Denmark almost all retirement goes through an early retirement program called post-employment wage (PEW) or old age pension (OAP). The PEW route is available from age 60 to age 64, and you earn eligibility for PEW through membership in an unemployment fund. Old age pension is a universal pay-as-you-go program that is available to everybody aged 65 and older. Since the focus is on how individuals changes retirement behavior when exposed to a health shock, it is crucial that the individual have a real choice to change their retirement behavior, therefore I ignore the DI pathway which is something you have to apply for.[21]

2.1 The PEW program

PEW is an early retirement program available from age 60 until 64/66[22]. You earn eligibility through membership in an unemployment insurance fund for at least 30 years.[23] The membership fee is 730 euro per year. The membership fee only covers about 40% of the total cost of the program–the rest is tax-financed. The yearly benefit amounts to 27,200 euro before means testing. The PEW benefit is means tested against almost all other individual income flows.

In the beginning of 1999 a reform of the PEW was announced which intended to reduce the use of the program. The reform affected individuals born after July 1, 1939 (who turned 60 in the second half of 1999). Throughout the paper I use the terms *before* and *after* 1999 to refer to the cohorts born before July 1 1939 (and not affected by the reform) and those born on July 1 1939 or later (who were affected by the reform). The reform changed economic incentives to retire in at least three important ways:

a) Before 1999, only life annuities that were paid out were means tested against PEW. After the reform, the means testing was extended, for those aged 60 and 61, to include all pensions (life annuities, fixed annuities and capital pensions) regardless if they were paid out or not.

---

[21] Retirement through the DI program is not a free choice since you have to fulfill certain criteria regarding health and work capacity in order to be awarded a DI pension. The DI program is also not a retirement program, since it is available from age 16. Later in this chapter I show that a very small share of the retirement from age 60 to age 67 go through DI.

[22] From 2004 to 2006 the age span was reduced from 60-66 to 60-64, because the OAP retirement age was reduced from 67 to 65 from 2004 to 2006.

[23] The seniority requirements have been increased several times since the program was introduced in 1979. Until 1999 the requirement was 25 years of membership.

b) Before 1999, the average benefit level (before means testing) was between 87 and 90% of the full unemployment benefit level if retiring before age 63 (depending on which age you retired at). After the reform, the PEW benefit rate was 91% of full benefits if retiring before age 62.[24]
c) After 1999, a person eligible for PEW earned a tax-free bonus of 1.700 euro every quarter from age 62 that he or she did not retire. If they never retired, the total tax-free bonus amounted to 20,400 euro at age 65.

Appendix C gives examples of the resulting means testing due to point a) for different types and sizes of the pension wealth. Point a) and c) both increased the incentive to postpone retirement until after age 61 relative to the period before the reform, whereas the effect of point b) is more ambiguous. This is because the benefit levels were about the same before and after the reform, but before the reform the incentive was to retire at age 63 instead of age 62, which gave an incentive to retire later before the reform. On the other hand, you could receive PEW for two extra years before the reform, and since the PEW benefit is higher than the OAP benefit, that change created an incentive to retire earlier before the reform. The measures of economic incentives that I use capture all these short (means testing in the PEW) and long term (the tax free bonus) effects of retiring today. Point b) and c) above clearly generated the largest economic incentives to postpone retirement on average.

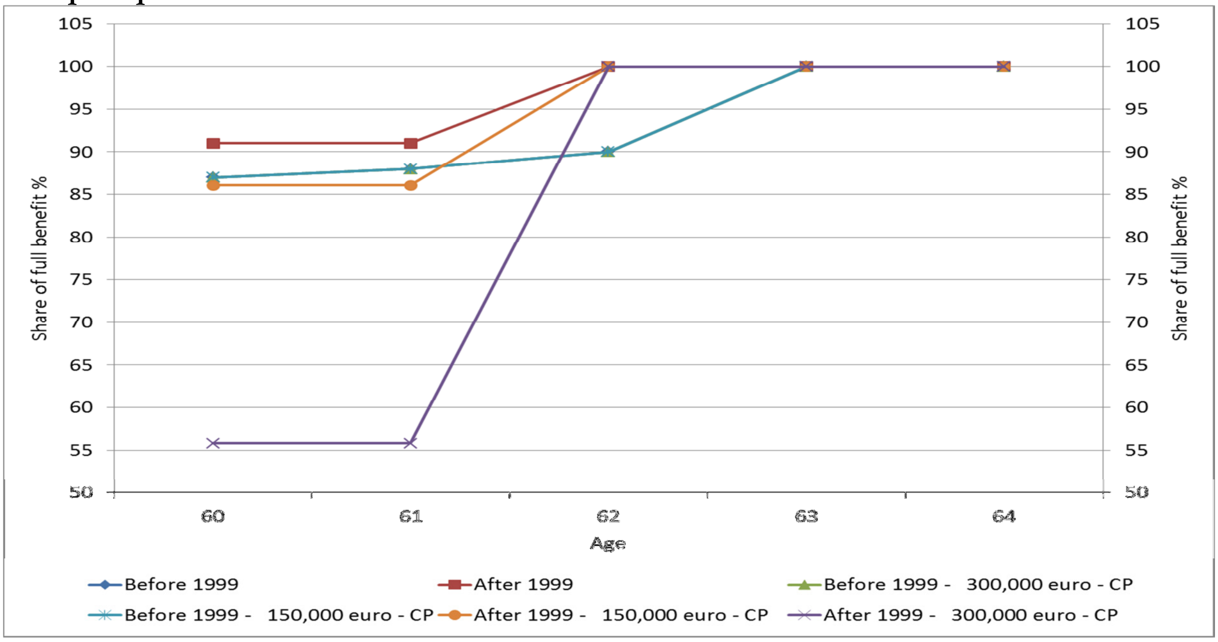
Figure 1 shows how different levels of pension wealth (capital pension) at different ages affect the benefit level (compared to the full benefit level) through means testing.

Before the reform, the benefit level was the same no matter the size of the capital pension, because capital pensions were not means tested. After the reform, the benefit level was reduced for all aged 60 and 61 no matter if the private pensions were paid out or not. After the reform, an individual with a capital pension amounting to 300,000 euro, and who retired before age 62, would have his/her benefit level reduced to 56% of the full benefit level, even when the individual postponed the payout of the capital pension to age 65.

---

[24] The reason why the incentive to retire was reduced from age 63 to age 62 was due to a reduction in the OAP retirement age from age 67 to age 65. Before 1999 a person could receive PEW for 7 years whereas a person born after the 1. Of July 1939 could only receive PEW for 5 years after the reform, and since the PEW benefit level is higher than the OAP benefit level, they were compensated by reducing the incentive to retire to age 62 instead of age 63.

Figure 1: Share of full PEW benefits before and after the reform split by retirement age and size of capital pension



The reform reduced economic incentives to retire for individuals holding capital or fixed annuity pensions no matter if they were paid out or not. The reform reduced economic incentives to retire for individuals with life annuity pensions that were not paid out when retirement started, while it had a modest effect on individuals without any private pension wealth, or individuals with life annuity pensions that were paid out when retirement started (see Appendix C for details).[25]

2.2 The OAP program

OAP is a universal tax financed demogrant pension. The "only" eligibility requirements are that you are a Danish citizen and have been living in in Denmark for at least 3 years. To receive the full amount you have to have lived in Denmark for 40 years between age 15 and age 65.[26] OAP is available from age 65/67.[27] OAP is a universal tax-financed pension consisting of a basic amount and a supplement. The basic amount and the supplement for a single individual are both 8,700 euro per year before means testing. For couples the base amount is the same, but the supplement is 4,350 euro per year. The supplement is means tested for all other incomes, whereas the basic amount is only means tested against income from work. Means testing for a single person starts at 8,150 euro per year for the

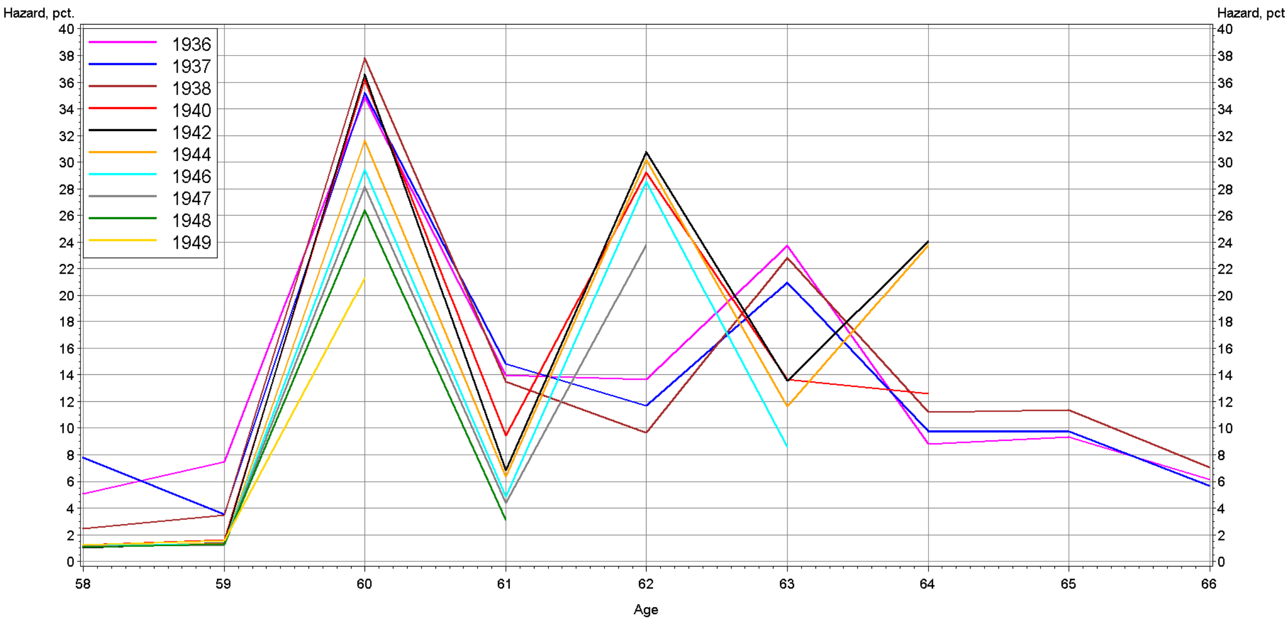
[25] Capital pensions, fixed annuities and life annuities can all be held as private or occupational pensions, but almost all life annuity pensions are occupational pensions.
[26] If you lived in Denmark between 3-40 years, you get a share of the full benefit corresponding to the share you lived in the country.
[27] From 2004 to 2006 the OAP retirement age was reduced from 67 to 65.

supplement and at 37,000 euro per year for the basic amount–the offset is 30 cent per euro.[28] Old age pensioner's with a maximum wealth of 10,400 euro (exclusive housing wealth) and a supplement income between 2,400 euro per year and 8,330 euro per year are eligible for a second supplement called "ældrecheck". If you earn less than 2,400 euro per year, you receive the full amount of 1,500 euro per year, and for every 59.3 euro your earn above the 2,400 euro, you loses 1% of the benefit. Couples are means tested together for all three elements (basic amount, supplement and "ældrecheck").

2.3 Observed retirement

This section illustrates the development in observed retirement hazard rates from 1994 to 2009 split by birth cohort. From figure 2 we see that the retirement hazard rate is very low until age 60, when PEW becomes eligible–only around 1-4% retires before age 60.

Figure 2: Retirement hazards by cohort (born 1936 to 1948), 1994-2009, All states



Note: The hazards include retirement to PEW and DI.

At age 60 the hazard rate increases to somewhere between 21% and 38% depending on which cohort we observe–the average hazard rate is lower at age 60 after the reform. Figure 2 strongly indicates that the incentive changes in the PEW reform affected retirement behavior–pre-reform the second spike is at age 63, whereas the second spike is at age 62 post-reform.

---

[28] For couples the means testing of the supplement starts at 16.400 euro per year and the offset is 15 cent per euro.

Figure 1b in Appendix A shows the hazard rates to PEW only. By comparing figure 1b and 2 it is clear that PEW represents almost all retirement from age 60 until OAP, which indicates how much the average retirement age is determined by the program. The next section describes the data and the estimation-sample.

3. Samples and descriptive statistics

I use two different samples in this paper. I need one sample to calculate income streams, health measures and survival probabilities 5 years out and I also need one sample for the estimation of the retirement models with the interaction between the health shock and economic incentives. The two samples spans different time periods and are selected from different criteria on observables. The next subsections describe the two samples and how I define the health shocks.

3.1 Sample for calculation of income streams, health and survival probabilities

In order to characterize economic incentives of a couple I need to calculate the expected present value of all future income streams from choosing different routes to retirement and choosing different points in time to retire. To do this, I use administrative register data for the period 1990 to 2008, on a 50% sample of the Danish population, which contains information on income from work, pension wealth, other wealth, labor market status, level of education, gender, ethnicity, spouses, marital status, diagnosis, eligibility for PEW and survival probabilities (by gender and age). This sample is primarily used to create summarized or historical variables for use in the estimation sample.

3.2 Estimation sample

To estimate the retirement models, I need to modify the sample in several ways in order to identify the interaction effect between health shock and economic incentives. Therefore, I apply the following selection criteria:

a) Only people who were in the sample at age 59 and age 60 and their spouses[29].
b) Only couples who were in the same relationship at age 59 and 60.
c) Only couples without diagnoses at age 59.
d) Only people who were not retired at age 59 (possible to be unemployed).
e) Only people who were eligible for PEW.
f) Only people who belonged to the sample during 1997 to 2004.
g) Only people who whose economic variables (income stream from work, income stream from pensions) were not outliers–I remove the 1'st and the 99th percentile of the observations.[30]

[29] Spouses include all types of couples–not only married couples.
[30] This is done because income and pension wealth distributions have very long tails–especially to the right (positive values).

All these criteria's reduce the sample to 120,379 panel observations. In the estimation section I use different controls to check the robustness of the marginal effects. I use education level (9 groups), gender, total income at age 53 (before the PEW reform), employment for the last 5 years, diagnosis 5 years back, sickness absence from work 5 years back, and days in hospital 5 years back.[31]

Table 1 shows different descriptive statistics for the final sample split by education and gender. The largest shares of the sample belongs to educational group 35 (Vocational education) and group 10 (Basic school). There are slightly more men than women in the sample because more women than men leave the labor force before age 59. The relatively low average retirement probability (16%) is due to the fact that almost nobody retires at age 59. Therefore, the 16% actually corresponds to 32% at age 60. The retirement probability clearly decreases when the educational level goes up and women retire with a higher probability than men (53% more often). The average option value is 63 and increases with educational level because those with higher educational levels earn higher incomes and hold more wealth (not shown here), which helps explains why their retirement probability is lower. Men's option value is on average 30% higher than women's, and this is one of the main reasons why men retire later than women.

Table 1: Descriptive statistics for the estimation sample, age 59 and 60

		Educational level–Higher number = higher level										Gender	
	All	10	20	25	35	40	50	60	65	70	Unknown	Men	Women
Observations	120,379	35,153	1,190	554	54,096	5,006	17,815	132	4,790	120	1,523	70,256	50,123
Share of all	100.0	29.2	1.0	0.5	44.9	4.2	14.8	0.1	4.0	0.1	1.3	58.4	41.6
Average													
Retirement probability	16	20	13	10	16	13	13	8	5	5	14	13	20
Option value	63	54	71	80	62	68	74	83	91	94	68	70	54
Total income at age 58 (1.000 kr.)	344	288	393	453	337	363	410	472	542	569	361	400	267
Employment 5 years back (share)	89	86	87	88	88	90	94	88	94	93	87	91	86
Diagnosis 5 years back (yes=1)	5.7	5.8	5.4	6.5	5.8	5.4	5.3	4.4	4.5	3.7	5.6	5.8	5.4
Sick absence 5 years back (share)	1.3	1.6	0.7	0.8	1.4	1.0	0.7	0.9	0.5	0.0	1.3	1.2	1.4
Days in hospital 5 years back	1.6	1.6	1.7	2.0	1.6	1.5	1.4	1.5	1.2	0.7	1.6	1.7	1.4

Note: Definitions of education level: 10=Basic school (8-10 grade), 20=General upper secondary school, 25=Vocational upper secondary school, 35=Vocational education, 40=Short-cycle higher education, 50=Medium-cycle higher education, 60=Bachelor, 65=Long-cycle higher education and 70=Research education.

Total income and employment 5 years back reflects the same pattern as the option value, in that both values rises with educational level and men have higher values, while the opposite is true for unemployment 5 years back. Relatively fewer people with higher education have been diagnosed in the

[31] I will comment on the choice of controls in section 4 in the identification section.

last 5 years compared to people with lower education, and the same pattern applies to days in hospital over the last 5 years. A slightly larger share of the men in the sample has been diagnosed in the last 5 years, and men have also spent slightly more days in hospital than women.

3.3 Definition of health shocks

I expect the general health level of an individual to be strongly correlated with individual preferences (which I do not observe), and I expect health measured in levels to be correlated with preferences, which makes health measured in levels endogenous. One way to avoid this endogeneity is by using health changes instead of health levels, and assume that the timing of the health change (a shock) is unknown to the individual. In order to observe the exact timing of the health change, I use information on yearly objective diagnoses from the National Patient Registry. From the National Patient Registry I can observe when a person was diagnosed and what diagnosis the person received.[32] In order to test the robustness of the interaction with economic incentives, I define the health change (health shock) in 3 different ways based on the objective diagnoses. The three different definitions are:

Health shock definition 1:

Any diagnosise received from t-1 to t is defined as being a health shock.

Health shock definition 2:

This definition was created by doctors from the branch organization "Insurance & Pensions" in Denmark. The doctors divided people into three groups according to their diagnoses: a) a very severe diagnosis, b) a less severe diagnosis c) no diagnosis. I collapse the doctors three groups into two groups–"very severe" and "not severe" (not severe and no diagnoses). Most of the very severe diagnoses according to this definition are different types of cancer and heart attacks.[33]

Health shock definition 3:

This definition is based on the ranking of the beta coefficients for 99 diagnoses in a survival model that controls for gender, age, year and ethnicity. The beta coefficient (from a linear probability model) measures the reduction in the five-year survival probability from specific diagnosis at a given point in time. I define an individual to have experienced a health shock if the size of the reduction in the five-

[32] The diagnosis is the dominant diagnosis at the end of the year. The diagnoses are divided into 99 different diagnoses, and you have to be hospitalized in order to receive a diagnosis.

[33] Cancer and heart attacks has often been used in the literature to define a health shock (see e.g. Conley and Thompson (2011).

year survival probability is larger than 5%. The beta coefficients on the "worst" diagnosis reduce the survival probability by 73.6% points, and at the 1'st percentile the reduction in the survival probability is 18.1% points.

Table 2 shows the share of the sample that receives a health shock according to the three health shock definitions. According to the first definition, HSPI and HSSI represent health shocks to one's partner and to one-self, respectively. In the same way, subscripts 2 and 3 refer to definitions two and three. According to health shock definition one 7% of the sample received a health shock–This applies to both HSP1 and HSS1. When I use the most severe definition (HSP2 and HSS2–definition two), only around 0.7% of the sample received a health shock from age 59 to age 60.

Table 2: Share of sample receiving a health shock from age 59 to 60, three definitions

	Observations	Mean
HSP - Health shock to partner - definition 1	55,507	0.070
HSS - Health shock to self - definition 1	55,507	0.072
HSP2 - Health shock to partner - definition 2	55,507	0.006
HSS2 - Health shock to self - definition 2	55,507	0.005
HSP3 - Health shock to partner - definition 3	55,507	0.030
HSS3 - Health shock to self - definition 3	55,507	0.032

Note: The table show the share of the estimation sample that received a health shock from age 59 to age 60. Definition 1 defines a health shock as all new diagnosis from t-1 to t, definition 2 defines a health shock as a new severe diagnose from t-1 to t, and definition 3 defines a health shock based on the effect on the survival rate 5 years out from a new diagnosis.

The third definition (HSP3 and HSS3–definition three)–which uses the effect of the diagnoses on the survival probability to define the health shock–affects 2.9% of the sample. A two-way frequency table of HSP1 and HSS1 (see table 1b in Appendix B), shows that in only 0.5% of the individuals experienced a health shock to self and a health shock to partner (according to health shock definition one) from age 59 to age 60. Therefore I estimate the models separately.

All three definitions avoid some of the measurement errors that self-reported health assessments gives rise to because they are based on objective diagnoses made by doctors. The weakness of using objective diagnoses is that they only apply to a small proportion of the sample and therefore they only capture part of the individual health dimension, whereas self-reported health assessments capture a broader definition of health.

4. Estimation and identification

To date, various methods have been developed and introduced to measure how health influences retirement/labor supply. These methods range from simple descriptive methods showing correlations between health level and retirement to very sophisticated structural models, that explicitly models investment in health and pension saving. I use a reduced form approach, and exploit unanticipated changes in economic incentives and the unknown timing of a health shock in order to identify how a health change affects the marginal economic incentive to retire.[34]

4.1 Empirical model

The main purpose is to assess how unhealthy individuals respond to a changed economic incentive to retire. To be able to do that, I define the following latent variable model formulated as a dif-in-dif model (DD):

$$(4.1) \quad y_{it}^{*} = \alpha + \beta_1 HS_{it} + \beta_2 OV_{it} + \beta_3 HS_{it} * OV_{it} + \beta_{4-20} X_{4-20} + \varepsilon_{it}, \qquad y_{it} = 1\left[y_{it}^{*} > 0\right]$$

where y* is an unobserved variable taking on the value 1 if a person retire and 0 otherwise. HS represents either a health shock to a partner (HSP) or a health shock to one-self (HSS).[35] OV is the option value which incorporates economic incentives to retire.[36] The error term ε_{it} is assumed to be independent of HS, OV and X and distributed symmetrically around 0. I further assume ε_{it} to be normal distributed, which allows us to formulate (4.1.) as a probit model, which ensures that the probabilities are strictly between 0 and 1.[37]

$$(4.2) \quad P(y_{it} = 1 \mid HS_{it}, OV_{it}, X_{it}) = \Phi(\alpha + \beta_1 HS_{it} + \beta_2 OV_{it} + \beta_3 HS_{it} * OV_{it} + \beta_{4-20} X_{it}^{4-20} + \varepsilon_{it})$$

I expect a health shock to one-self (HS=HSS) to increase the probability to retire ($\beta_1 = +$), whereas the effect of a health shock to one's partner (HS=HSP) is more ambiguous ($\beta_1 = \pm$). An increase in OV,

[34] Incorporating health problems into a standard DP retirement model is complex, because health is likely to affect the timing of retirement in many ways. Poorer health often has a negative impact on productivity, can reduce earnings and affect preferences, affect the utility of consumption and leisure, and influences individuals' life expectancy and hence the number of years available to choose between work and retirement (Deschryvere, 2005).

[35] I focus on either a health shock to the partner or a health shock to self, to keep the model as simple as possible. Incorporating both a health shock to self and to partner, does not change the main results, and the interaction term HSP*HSS is not significant.

[36] See Appendix A, for a detailed description on how it is calculated.

[37] The Probit formulation allows the marginal effect to be non-constant across values of the independent variables, which has shown to be an important feature in retirement modeling.

indicates that the present value from retiring today is further away from the maximum present value of retiring at all further ages. Therefore I expect β_2 to be negative. The sign of the interaction term (β_3) measures how individuals will change their response to economic incentives, if they experiences a health shock to them self (HS*OV=HSS*OV) or to their partner (HS*OV=HSP*OV). Direct interpretation of the sign or the level of significance for an interaction term is not valid in non-linear models (see Dowd et al., 2011). Therefore I calculate the marginal effects (dy/dx in STATA) and their corresponding significance levels in order to interpret the effect of the interaction between the health shock and economic incentives. The expected effect described in the text is to be understood as the marginal effect from dy/dx and not the coefficient from the probit model. I expect β_3 to be positive if the health shock is to self and either positive or negative if it is to one's partner, depending on if the added worker effect (negative β_3) or the care/leisure effect (positive β_3) dominates. A positive β_3 indicates a reduction in the response to economic incentives.

4.2 Identification

To identify the causal marginal effect of the interaction term, health and economic incentives have to be exogenous, which will ensure that they are not correlated with ε_{it}. Since I expect health (measured in levels) and economic incentives to be correlated with preferences for leisure, saving and health investment, which I do not observe, I expect health (measured in levels) and economic incentives to be correlated with the error term. To circumvent the fact that health and unobserved preferences are correlated, I define health as a health shock and assume that the individual does not know the exact ***timing*** of the health shock. Even though some individuals might expect a health shock with a higher probability than others, because of their life style (smoking, drinking, weight etc.), they will not know the exact timing of the health shock. At the same time, I exploit the variation in economic incentives that the 1999 PEW reform created. Since the PEW reform was not announced until spring 1999, and the reform was implemented by the July 1 1999, I assume the changed economic incentive to be an unanticipated change which is uncorrelated with unobserved preferences (corr(ε_{it}|HS*OV)=0).

Overall, the variation in OV and health shock (HS) in equation (4.2) stems from 3 different sources:

a) Observed variation between individuals due to the unanticipated timing of the health shock and the incorporation of the PEW reform into the OV calculation.
b) Unobserved variation between individuals due to heterogeneity (ε)
c) Observed variation between individuals due to heterogeneity (X)

To estimate the causal effect, I need to minimize the variation from the unobserved variation (ε). I do that by including as much variation from the PEW reform in the calculation of the OV as possible and by controlling for important background information such as gender, year, education, permanent income at age 53 (before reform) and historical health variables.

Permanent income before the reform captures the economic situation before the reform and is therefore not influenced by the changed incentives in the reform. The historical health variables helps to control for health level, which partly reflects life time investment in health (preferences for health). The three different historical health measures (diagnoses -last 5 years; sickness absence from work -last 5 years; days in hospital -last 5 years) capture different elements of the health level. Diagnoses during the last 5 years and days in hospital during the last 5 years have the advantage of being objectively assessed by doctors, but they only span a smaller part of the sample. Sickness from work during the last 5 years capture a brother definition of the health level but is also more a choice variable, and may therefore also reflect preferences for leisure to some extent.

5. Results

In this section I estimate different retirement models each of which shows the interaction of a health shock with economic incentives in order to see, if health deterioration reduces the responsiveness to economic incentives. First I illustrate the robustness of the interaction effect when important controls are added, and then I estimate the model with a full set of controls with 3 different definitions of the health shock to see, how different definitions influences the marginal effect of the interaction term. Then I estimate models split by gender, to see if women and men react differently (asymmetric) to their spouses' health shocks. Finally I estimate observed, predicted, and simulated retirement probabilities for a retirement model with the health shock defined by the reduction in the survival probability 5 years out from 99 diagnoses. In all models I estimate both the effects of a health shock to self and to partner, and all marginal effects are estimated as dy/dx in STATA's MARGINS command.[38]

5.1 Retirement models with different controls

First I estimate a retirement model with different controls to see how much of the effect from the economic incentive and the interaction term that is due to heterogeneity. Table 3 shows the retirement models estimated for the third health shock definition (defined by the reduction in the survival probability 5 years out from 99 diagnoses).

Models 1-4 show the results for a health shock to one-self, and models 5-8 show the result for a health shock to one's partner. The marginal effect on OV is negative and very significant across all 8 models, indicating that the bigger the difference between the present value of all income streams from retiring today and retiring at the point that creates the biggest present value (the max), the less likely it is that a person will retire. HS3 measures the direct effect of getting a health shock from t-1 to t. The direct effect is highly significant across all 8 models.

The marginal effect on the interaction term, which measures the reduction to economic incentives due to a health shock, is positive and highly significant across all 8 models, although the size of the marginal effect is moderate. The marginal effect on the interaction term decreases across the 8 models as more controls are added, and the size of the marginal effect varies between 0.41 and 0.13. The results indicates, that an individual reacts a bit stronger to a health shock to one's partner as to a health shock to one-self, when a health shock is defined by health shock definition three (defined by the reduction in

[38] Standard errors are calculated with the user written command ESTEFF.

the survival probability 5 years out from 99 diagnoses). The care/leisure effect seems to dominate the AWE when the health shock happens to one's partner.

Table 3: Retirement model, OV interacting with HSS3 and HSP3, Different controls

	HSP3				HSS3			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.
OV at HS = 0	-4.9113***	-4.9883***	-4.9943***	-4.9946***	-4.9084***	-4.9804***	-4.9852***	-4.9845***
	(0.0399)	(0.0416)	(0.0469)	(0.0470)	(0.0398)	(0.0416)	(0.0469)	(0.0469)
OV at HS = 1	-4.5018***	-4.6737***	-4.7675***	-4.7707***	-4.5472***	-4.6875***	-4.8554***	-4.8510***
	(0.3016)	(0.2954)	(0.2819)	(0.2816)	(0.3225)	(0.3111)	(0.2918)	(0.2927)
Difference	0.4095	0.3146	0.2267	0.2238	0.3612	0.2929	0.1298	0.13347
P > |z| difference	0.0095	0.0070	0.0047	0.0047	0.0039	0.0032	0.0013	0.0013
HS3	0.1019***	0.0976***	0.0945***	0.0943***	0.1213***	0.1171***	0.1102***	0.1109***
	(0.0117)	(0.0114)	(0.0111)	(0.0111)	(0.0115)	(0.0111)	(0.0107)	(0.0107)
Gender		x	x	x		x	x	x
Year		x	x	x		x	x	x
Education		x	x	x		x	x	x
Employment last 5 years			x	x			x	x
Income in t-5			x	x			x	x
Diagnosis last 5 years			x	x			x	x
Sickness 5 last years			x	x			x	x
Days in hospital last 5 years				x				x
Observations	120,362	120,362	119,450	119,450	120,362	120,362	119,450	119,450
Pseudo R2	0.225	0.232	0.238	0.238	0.224	0.231	0.238	0.238

Note: OV is the option value measurement. HSS3 is a health shock to self according to definition 3 (defined by reduction in survival probability 5 years out from a new diagnosis). HSP3 is a health shock to partner according to definition 3.

We can conclude that economic incentives have a very strong effect on retirement, and that those who experience a health shock are significantly less responsive to economic incentives (independent of the number of controls we add to the specification), but the magnitude of the reduction is moderate.

5.2 Retirement model with different health shock definitions

This section estimates a retirement model with the full set of controls from table 3, where OV is interacted with all three definitions of a health shock. Subscripts 1, 2 and 3 refer to the three definitions of the health shock, respectively (1 = all new diagnoses; 2 = only very severe new diagnoses; 3 = only diagnoses that give rise to more than a 5% point reduction in the 5 year survival rate. The results are seen in table 4.

Again, the marginal effect on OV is highly significant and stable across the 6 model specifications, and the health shock level effect is again significant. 5 out of the 6 marginal effects on the interaction terms are positive and significant. If we first compare the marginal effect on the health shock to self (HSS) and the health shock to partner (HSP) for the three health shock definitions, then we see, that the magnitude is a bit stronger if the health shock happens to one's partner instead of one self, when the health shock is defined by definition one and 3, whereas the opposite is the case for definition two. For definition two there is an AWE effect when the health shock occurs to one's partner.

Table 4: Retirement model, OV interacting with 3 different health shocks

	HSP1	HSS1	HSP2	HSS2	HSP3	HSS3
	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14
	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.
OV at HS = 0	-4.9815***	-4.9704***	-5.0098***	-5.0121***	-4.9946***	-4.9845***
	(0.0470)	(0.0477)	(0.0466)	(0.0466)	(0.0470)	(0.0469)
OV at HS = 1	-4.5028***	-4.5786***	-5.1790***	-4.9275***	-4.7707***	-4.8510***
	(0.2308)	(0.1945)	(0.5452)	(0.6616)	(0.2816)	(0.2927)
Difference	0.4787	0.3918	-0.1692	0.0845	0.2239	0.1335
P > |z| difference	0.0010	0.0006	0.0556	0.0922	0.0047	0.0013
HSP	0.0860***		0.1054***		0.0943***	
	(0.0080)		(0.0217)		(0.0111)	
HSS		0.0863***		0.1083***		0.1109***
		(0.0071)		(0.0263)		(0.0107)
All controls	x	x	x	x	x	x
Observations	119,450	119,450	119,450	119,450	119,450	119,450
Pseudo R2	0.239	0.239	0.237	0.237	0.238	0.238

Note: OV is the option value measurement. HSS represents a health shock to self and HSP represents a health shock to partner. The subscript represents which health shock definition that is used. Definition 1 defines a health shock as all new diagnosis from t-1 to t, definition 2 defines a health shock as a new severe diagnose from t-1 to t, and definition 3 defines a health shock based on the effect on the survival rate 5 years out from a new diagnosis.

If we compare the magnitude of the marginal effect on the interaction term from the 6 models, then we see a huge variation in the size of the effect. In the most general definition of a health shock (definition one), the reduction in the economic response is 3 times greater than the reduction caused by the most severe definition of a health shock (definition two), while it is only 2 times as great compared to definition three.

We can conclude: 1) that the definition of the health shock is important for understanding the size of the effect of the interaction term, and 2) that the marginal reduction in the response to economic incentives is a bit stronger when the health shock happens to one's partner than to one self.

5.3 Retirement model split by gender

Some studies (see e.g. Coile, 2004b; Gustman and Steinmeier, 2000 and 2004; McGeary, 2009) have shown that there exists asymmetry in the reaction of men and women to a health shock to themselves and to their partners–normally measured directly without interacting health with economic incentives.

Table 5 shows the effect on the interaction term from a health shock to one-self and to one's partner split by gender and estimates using health shock definition three (defined by the reduction in the survival probability 5 years out from 99 diagnoses).

Table 5: Retirement model, OV interacting with HSS3 and HSP3 split by gender

	Male	Female	Male	Female
	Model 15	Model 16	Model 17	Model 18
	margins/Std.	margins/Std.	margins/Std.	margins/Std.
	HSP3	HSP3	HSS3	HSS3
OV at HS = 0	-4.0316***	-6.8117***	-4.0111***	-6.8184***
	(0.0569)	(0.0837)	(0.0571)	(0.0831)
OV at HS = 1	-3.7328***	-6.4727***	-3.9786***	-6.3707***
	(0.3985)	(0.4101)	(0.3410)	(0.5764)
Difference	0.2988	0.3390	0.0325	0.4478
P > |z| difference	0.0486	0.0906	0.0043	0.0666
HS3	0.0805***	0.0874***	0.1001***	0.1344***
	(0.0147)	(0.0161)	(0.0126)	(0.0224)
All controls	x	x	x	x
Observations	69,683	49,767	69,683	49,767
Pseudo R2	0.225	0.255	0.226	0.256

Note: OV is the option value measurement. HSS3 is a health shock to self according to definition 3 (defined by reduction in survival probability 5 years out from a new diagnosis). HSP3 is a health shock to partner according to definition 3.

First we see that women on average respond stronger to economic incentives than women, this might be because that the marginal response is decreasing in the size of the OV (not shown here), and women on average have lower OV's than men. Table 5 also shows that the reduction in the response to economic incentives due to a health shock to one's partner is about the same for males and females. This finding changes when we look at a health shock to one-self, where the marginal reduction in the economic response of the female is 14 times larger than men's reduction.

This suggests that the care/leisure effect dominates for both males and females when their partners experience a health shock, and that women are much more responsive to a health shock to self than men are.

5.4 Retirement model split by life annuity payout

This section illustrates how the marginal response to the interaction between the health shock and economic incentives varies with the size of the expected life annuity payout at age 59. [39] Table 2b in Appendix B shows descriptive statistics for the four quartiles of the expected payouts. From table 2b we see that the expected life annuity payout is negatively correlated with the OV measure and positively correlated with the income at age 53. The table also show that when we move upwards from quartile 1 to 4 we see that: a) weeks employed goes up, b) average time diagnosed the last 5 years decreases, c) sickness absence from work the last 5 years decreases, and d) days spent in hospital decreases. The share of each quartile that received a health shock (defined by definition three) is not very different– only quartile 2 have a somewhat larger share than the other 3 quartiles. One explanation why the share is not that different might be because I condition on people having no diagnosis at age 59, which might have removed some of the weakest individuals from group 1.

Table 6 show extracts from the estimated retirement models split by life annuity quartiles and a health shock (defined by the reduction in the survival probability 5 years out from 99 diagnoses) to one-self (HSS3) or to ones-partner (HSP3). The full output is seen in table 3b in Appendix B.

The marginal response to OV decreases as the quartiles on of the expected pension payouts rises–it falls from around -7.1 to -3.5 regardless whether HSP3 or HSS3 is used as the health shock. The interaction term on the health shock to one's partner (HSP3 * OV) is very different across the four quartiles–it is very large in quartile 1 then it falls in quartile 2 and 3, and rises again in quartile 4. The large variation in the interaction term of OV*HSP3 might be because of the large variation in the OV and the fact that the interaction term is only significant in quartile 2, where the effect is negative (AWE). The size in the marginal effect on the interaction term on the health shock to one-self (HSS3 * OV) also varies a lot across the four quartiles, but is also insignificant in three out of the four quartiles, and now there is an AWE in quartile 4. It is difficult to interpret the effect of the size of the interaction

[39] When calculating the expected payouts I assume that the entire pension wealth is a life annuity.

terms across these different groups without estimating retirement probabilities, because the size of OV and demographic characteristic may all have an effect on retirement probabilities.

Table 6: Retirement model OV interacted with HSS3 and HSP3 split by life annuity

	HSP3				HSS3			
	1. quantile	2. quantile	3. quantile	4. quantile	1. quantile	2. quantile	3. quantile	4. quantile
	Model 19	Model 20	Model 21	Model 22	Model 23	Model 24	Model 25	Model 26
	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.
OV at HS = 0	-7.1306***	-5.8509***	-4.2325***	-3.5332***	-7.1124***	-5.8601***	-4.2188***	-3.5125***
	(0.0952)	(0.0908)	(0.0947)	(0.0799)	(0.0945)	(0.0905)	(0.0945)	(0.0800)
OV at HS = 1	-4.6666***	-6.1815***	-4.1715***	-3.3365***	-5.0161***	-5.3563***	-3.5757***	-4.1536***
	(0.9088)	(0.5872)	(0.4788)	(0.5285)	(0.9747)	(0.7148)	(0.5821)	(0.4125)
Difference	2.4639	-0.3306	0.0610	0.1967	2.0963	0.5039	0.6431	-0.6411
P > |z| difference	0.8171	0.0342	0.1331	0.1113	0.7220	0.0724	0.1836	0.0208
HS3	0.1492***	0.1067***	0.0559**	0.0714***	0.1770***	0.1148***	0.1076***	0.0717***
	(0.0294)	(0.0235)	(0.0180)	(0.0202)	(0.0280)	(0.0232)	(0.0227)	(0.0162)

Note: OV is the option value measurement. HSS3 is a health shock to self according to definition 3 (defined by reduction in survival probability 5 years out from a new diagnosis). HSP3 is a health shock to partner according to definition 3. The quartiles are split by the size of the expected yearly life annuity payout from all pensions.

This pattern is clearly different from the over-all average marginal effect on the two interaction terms in table 4–now only quartile 1 exhibits the same relative magnitude between HSP3 and HSS3. One explanation for this pattern could be that those in quartile 1 hardly are means-tested and therefore the cost of responding to own health or the partner's health is not very high. In the second and third quartile on the other hand almost all individuals are being means-tested and their partners probably also earn more, so the cost if both spouses retire increases strongly. In the last quartile the pattern are opposite to the second and third quartile. One explanation could be that a share of those in quartile four on average can expect almost full means-testing in the PEW, and therefore they find it too expensive to retire because of a health shock. It is important to notice that the combined effect of the marginal effects on the OV, the health shock and the interaction term and the level of the OV, might still result in lower retirement probabilities in quartile four for someone who experience a health shock, because of the stronger marginal effect on the OV and the higher level of the OV.

From table 6 we can conclude that the interaction effect (HS * OV) varies over the four life annuity quartiles, and that the variation is different depending on if we study a health shock to one-self or to one's partner.

5.5 Predicted retirement probabilities

To illustrate the combined effect of the marginal response to OV, the marginal response to a health shock, the marginal response to the interaction term (HS * OV), the size of OV and whether you received a health shock or not, I predict the retirement probabilities for each quartile of the expected life annuity payout, to show, that a health shock affects individuals differently in the four quartiles. I also split the probabilities by "before" and "after" the reform of the PEW in 1999, to show, how the changed economic incentives from the reform affected the retirement probability differently in the four life annuity groups. In all the predictions I calculate the retirement probabilities at the average of the observed variables (education level, employment the last 5 years, income at age 53, years with diagnoses the last 5 years, work absence the last 5 years and days in hospital in the last 5 years) [40] in each life annuity quartile to eliminate differences on observables between the before and after sample and those with and without a health shock. The difference in the predicted retirement probabilities between "before" and "after" for each life annuity quartile hence reflect differences due to changed economic incentives (measured by OV) and a health shock (with fixed observables).

5.5.1 Predicted retirement probabilities split by life annuity groups, health shock, and before and after the reform in 1999

The PEW reform in 1999 increased the means testing for private pension payouts for individuals who retire before age 62. I expect that the increased means testing reduced retirement relatively more in higher quartiles of the expected life annuity payout, because they would lose a relatively larger share of their PEW benefit if they retired. I expect those who experienced a health shock to reduce their response to economic incentives by less, the higher the quartile, because the relative decrease in income is larger for individuals with higher incomes[41], because of the flat rate structure of the PEW benefit. I start out by comparing the observed and predicted retirement probabilities.

Table 4b in Appendix B show the observed retirement probabilities, the average life annuity payout and health shock frequencies split by life annuity quartile, health shock, and before and after the reform in 1999. From table 4b we see the following: a) The frequency of a health shock is about the same level in quartile 1, 3 and 4, while it is a bit higher in quartile 2, b) The average life annuity payout increases strongly over the four quartiles–from 7,897 kr. In quartile 1 to 168,011 kr. in quartile 4 and c) The

[40] I calculate the probabilities as if everybody was a man and I choose the most frequent educational level for each life annuity group.

[41] Income and expected life annuity payout is strongly correlated.

observed retirement probability decreases over quartiles–from 41.5% in quartile 1 down to 24.6% in quartile 4.

Table 5b in Appendix B show the predicted retirement probabilities split by life annuity quartile, health shock, and before and after the reform in 1999. From table 5b we see that the average predicted retirement probability is almost the same in the model with a health shock to one-self and the model with a health shock to one's partner, and the predicted retirements also decreases over the four quartiles as we saw with the observed probabilities–from 36% in quartile 1 down to 22% in quartile 4.[42]

To assess whether the PEW reform decreased retirement relatively more in the higher quartiles, I calculate the difference between retirement probabilities before and after the reform for each of the four quartiles for the observed retirement probabilities (from table 4b) and for the two predicted probabilities (from table 5b–model with HSS and HSP). The results are seen table 7.

Table 7: Change in observed and predicted retirement probabilities before and after the reform in 1999, split by life annuity group and health shock

Life annuity quantile	Health shock	Difference in retirement probability before and after reform %			Difference in retirement probability before and after reform–indexed with difference in quartile 1 = 100		
		Observed	Predicted-HSS3	Predicted-HSP3	Observed	HSS3	HSP3
1		5.6	31.2	31.3	100.0	100.0	100.0
2		0.5	32.8	33.2	94.9	101.6	101.9
3		-12.5	18.1	17.7	81.8	86.9	86.4
4		-25.6	10.3	10.8	68.7	79.0	79.5
1	No	6.0	32.6	32.5	100.0	100.0	100.0
	Yes	-4.3	7.7	7.7	100.0	100.0	100.0
2	No	0.7	33.8	34.0	94.7	101.3	101.5
	Yes	-3.3	19.1	20.5	101.0	111.4	112.8
3	No	-12.1	18.5	18.1	81.9	85.9	85.6
	Yes	-25.2	8.7	15.1	79.1	101.0	107.5
4	No	-25.5	10.8	10.8	68.4	78.3	78.3
	Yes	-27.3	4.1	3.6	76.9	96.4	96.0

Note: HSS3 is a health shock to self according to definition 3 (defined by reduction in survival probability 5 years out from a new diagnosis). HSP3 is a health shock to partner according to definition 3. The quartiles are split by the size of the expected yearly life annuity payout from all pensions.

If we first look at the column with the difference between observed retirement probabilities before and after the reform over the four quartiles, then we see that retirement on average was 5.6% higher after the reform in the first quartile. In the second quartile the probability was about the same (difference =

[42] One difference to notice is that the predicted retirement probability is higher after the reform in all four life annuity quartiles–I expect this partly to be due to the control for observables.

0.5%), in the third quartile the average retirement probability was 12.5% larger before the reform, and in the last quartile the retirement probability was 25.6% larger before the reform. This clearly indicates that the reform affected those in the higher quartiles relatively more than those in the lower quartile. It makes sense that retirement was not reduced by much before the third and fourth quartile, because the first 24.100 kr. of income was exempted from means testing,[43] and from table 4b we see that the second quartile starts at life annuity payouts of 14,508 and ends at 32,928 kr. So a large share of those in the second quartile would receive life annuity payouts that were too small to be means tested against PEW benefits.

The observed retirement probabilities do not control for different characteristics between the before and after sample, and they are not calculated based on the economic incentive from the reform, so I expect them to overestimate the effect of the reform. Therefore I also compare the predicted retirement probabilities from table 5b, before and after the reform for each quartile. The predicted probabilities are calculated for the average individual in each quartile and they are based on economic incentives (measured by the option value), which directly incorporates the economic incentive of the PEW reform. The predicted retirement probabilities are higher in all quartiles after the reform (≠observed probabilities), but the pattern in the relative differences between before and after the reform over the four quartiles is the same as we saw for the observed retirement probabilities. The reason why the retirement probabilities are higher in all quartiles after the reform is that the OV was lower after the reform, because there was a massive increase in occupational labor market savings from the mid 1990's (both number of savers and saving rate), which gave those who retired post reform a much larger supplement to the PEW, even though they were means tested harder.[44]

The last column of table 7 makes this point more clear, by indexing the difference between before and after the reform within each quartile, with the difference between before and after the reform in quartile 1 set to 100. The table show that the index decreases for both the observed and the predicted retirement probabilities in the second and the third quartile.

---

[43] The deductible amount was reduced in 2002 to 18.000 kr. and again in 2003 to 11.900 kr.–after 2003 the level remained at 11.900 kr.

[44] It is important to remember that a lot of this expanded savings were done in occupational life annuities that was not affected by the changed means testing in the 1999 reform, if the life annuity was paid out at retirement.

Based on these findings I conclude that the changed economic incentives in the PEW reform reduced retirement relatively more for those who could expect an increased means testing due to the reform.

The lower part of table 7 again show the difference between the retirement probabilities for the observed and predicted model before and after the reform for each of the four quartiles, but this time also split by the health shock, to see if those who received a health shock reacted in the same way to the reform as those who did not (Figure 3b-6b in Appendix B illustrates the predicted retirement probabilities split by life annuity groups, health shock, and before and after the reform).

If we first look at the index column for the observed retirement probabilities in table 7, then we see that the pattern is almost the same for those who received a health shock and those who did not. The observed probabilities do not control for differences on observables between the before and after group or between those with and without a health shock. Therefore I estimate the predicted probabilities which control for observables by giving everybody in each quartile the same characteristic (average in each quartile) on the X's (not OV!). From the index column for the predicted retirement models we now see that the index do not follow the same pattern, for those with and without a health shock–the relative reduction in the retirement probability over the four quartiles is almost not reduced compared to the first quartile for those who received a health shock (or received a health shock to partner). This indicates that the reduced economic incentive to retire that the PEW reform imposed did not succeed in reducing retirement relatively as much for those in bad health as for those in god health. Instead it primarily made them worse off economically.

6. Conclusions

One way of motivating people to retire later is by increasing economic incentives to keep working. Knowing whether this strategy will work for the less healthy is essential for policy makers, as the demographic trend is projected to increase over the coming three to four decades. To answer this question I use rich register data from Denmark to examine how a health shock to one-self and to one's partner influences the response to economic incentives (of the couple) to retire.

One important challenge when trying to get unbiased estimates of the effect of health on retirement is the endogenous character of the relationship between health and retirement. To circumvent this endogeneity, I exploit a 1999 reform of the early retirement program "Post Employment Wage" (*efterløn*, PEW) and the arguably unanticipated timing of health shocks.

The estimations show that the marginal effect on economic incentives (option value) is negative and very significant across all models, indicating that the bigger the difference between the present value of all income streams from retiring today and retiring at the point that creates the biggest present value (the max), the less likely it is that a person will retire.

The marginal effect on the interaction term between economic incentives and a health shock (to one-self or one's partner), which measures the reduction to economic incentives due to a health shock, is positive and highly significant across all models, although the size of the marginal effect is moderate. The marginal effect on the interaction term decreases across models when more controls are added.

A health shock to one-self or to one's partner reduces the response to the economic incentive by about the same amount. The care/leisure effect therefore seems to dominate the added-worker effect (AWE) when the health shock happens to one's partner.

The analysis show that the definition of the health shock is important for the size of the effect of the interaction term (OV * HS), and that the marginal reaction to a health shock to one-self and to one's partner is about the same across health shock definitions.

Comparing men and women's respective responses to economic incentives and health shocks show that women on average respond stronger to economic incentives than men, which might be because the marginal response is decreasing in the size of the economic incentive. The reduction in the response to economic incentives due to a health shock to one's partner is about the same for males and females, but women reduce their response to the economic incentive by 14 times as much as men when they experience a health shock to them self. This suggests that the care/leisure effect dominates the substitution effect for both males and females.

Finally I illustrate how the changed economic incentives in the PEW reform in 1999 affected individuals with different expected pension payouts differently. The predicted retirement probabilities show that those in the two lowest quartiles of the life annuity payout changed their retirement probabilities relatively less than those in the two largest quartiles. This illustrates that the economic incentives in the PEW reduced incentives to retire more for those with larger pension wealth. When I split these results by whether an individual received a health shock or not, then there is almost no difference in the relative change in the retirement probability across the four quartiles of the expected life annuity payout. This indicates that those in bad health had their incomes reduced due to the reform, because they were not able to keep on working because of the health shock they experienced.

Appendix

Appendix A: Economic incentives

1a Couples Option Value

I jointly model economic incentives of the husband and wife. Economic incentives are calculated in an Option Value model (OV-model) following Stock and Wise (1990). Joint behavior is modeled by incorporating joint means testing and allowing in the utility function the value of income for spouses to depend upon own and spousal retirement date. In a full maximum likelihood version of a couple's option value model, the couples are allowed to retire at any point in time. To simplify the calculation I assume that the spouse retires at the first possible retirement age (60) if the spouse is a women, and at the last possible date (65) if the spouse is a man.

2a The Option Value calculation

The OV compares the difference between the present discounted value of all future income streams from retiring today and retiring at r^{max}. Equation (1a) shows the present discounted value of all future income streams from retiring today:[45]

$$\text{(1a)} \qquad V_t(R) = \sum_{s=t}^{R-1} \beta^{s-t} \pi(s \mid t) E_t(Y_s^{\gamma}) + \sum_{s=R}^{T} \beta^{s-t} \pi(s \mid t) E_t\left[k * B_s(R)\right]^{\gamma}$$

t is current age, R is age of retirement, T is maximum possible age, β is the discount factor, π is the probability of surviving to age *s,* conditional on surviving to age t, Y is the income stream from working, $B_s(R)$ is the pension benefit at age *s* conditional on retirement at age R, k measures the relative value of income during retirement and γ is a risk aversion parameter (measures marginal utility of money). **Since equation (1a) is a semi-structural model, I will only use the semi- structural parameters β, k and γ as fit parameters**.

Equation (1a) is calculated for all possible retirement routes of the individual.[46] Then these present values are subtracted from the route that generates the maximum present discounted value, and the

---

[45]We incorporate a survival probability in our model–this was not in the original Stock and Wise (1990a) model. In that model you died at a fixed age.

[46]We allow for a 5 age span between husband and wife.

difference is the OV–see equation (2a). I use the economic streams of the household in order to maximize the utility function:

(2a) $$OV_t(R^*) = \sum_{s=t}^{R^*-1} \beta^{s-t} \pi(s|t) E_t(Y_s^\gamma) + \sum_{s=R^*}^{T} \beta^{s-t} \pi(s|t) E_t\left[k * B_s(R^*)\right]^\gamma - \sum_{s=t}^{T} \beta^{s-t} \pi(s|t) E_t\left[k * B_s(R)\right]^\gamma$$

R* is the retirement age that generates the maximum present income stream. According to the model, I expect that an increasing OV will generate a decreasing probability to retire and vice versa. The OV measurement and the health shock will be part of the X-vector in the final model, and the optimal parameter values are found by grid search over values of γ *and* k. β is assumed fixed at 0.97[47]. The grid search for parameters k and γ for β fixed at 0.97, gives k=1.75 and γ=0.3.

[47] The value of 0.97 is also used in other estimations of the OV-model (se e.g. Gruber and Wise, 2004).

Appendix B: Descriptive statistics and results

Figure 1b: Retirement hazard rate by cohort, born 1936-1948, 1996-2008, PEW

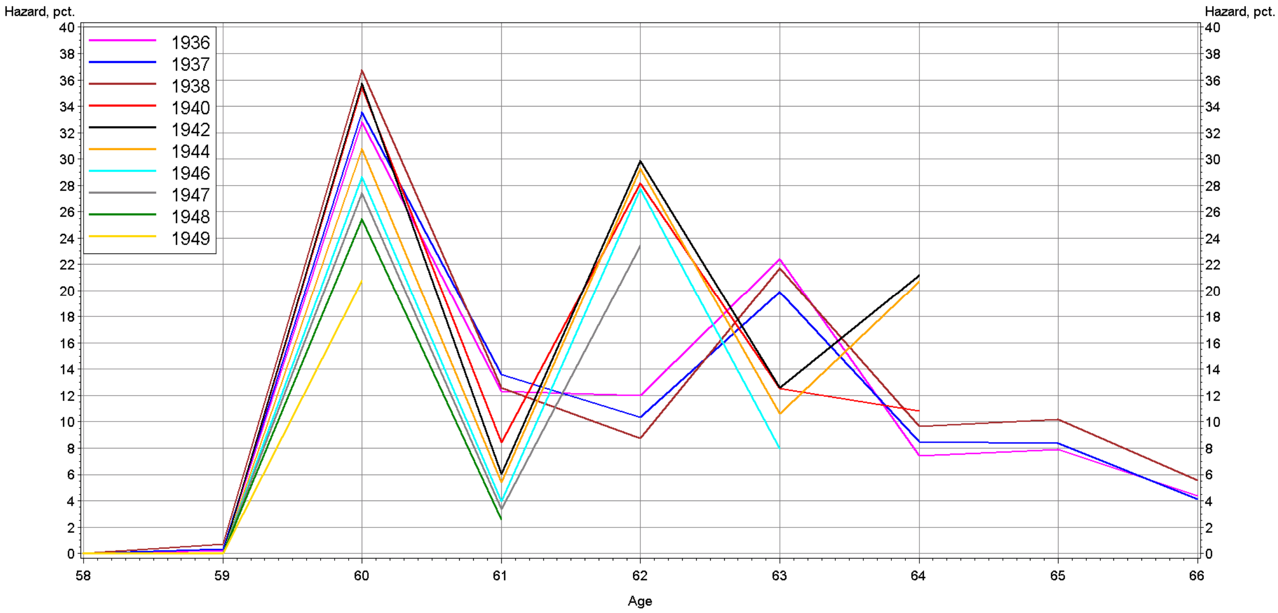


Table 1b: Health shock to self and to partner defined by health shock definition 1

Health shock self	Health shock partner	Observations	Share %
No	No	47,933	86.35
No	Yes	3,593	6.47
Yes	No	3,671	6.61
Yes	Yes	310	0.56

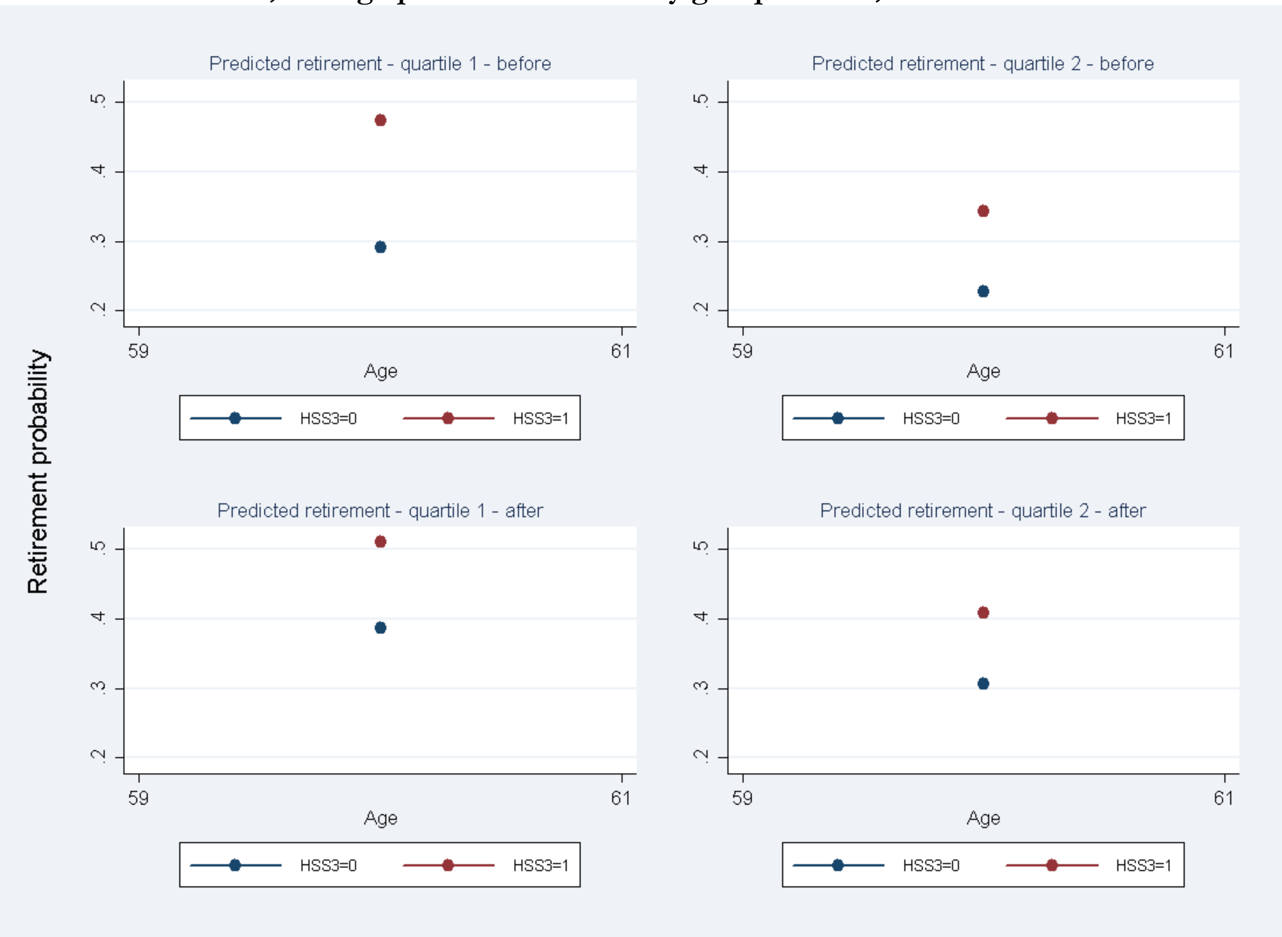
Table 2b: Descriptive statistics split by life annuity quartile

Quantile	Observations	Retirement probability	OV	HSS3	HSP3	Employment weeks 5 years back	Income at age 53, kr.	Years with diagnoses–last 5 years	Sickness absence in weeks last 5 years	Days in hospital last 5 years
1	14,405	38.9	33.2	0.030	0.030	215	253,841	0.286	4.4	2.0
2	12,286	35.2	37.1	0.036	0.031	235	298,231	0.270	3.7	1.6
3	11,190	29.2	45.6	0.032	0.031	245	349,242	0.252	2.5	1.4
4	11,465	26.1	53.2	0.032	0.027	251	450,445	0.229	1.6	1.3

Table 3b: Retirement model, OV interacted with HSS3 and HSP3 split by life annuity

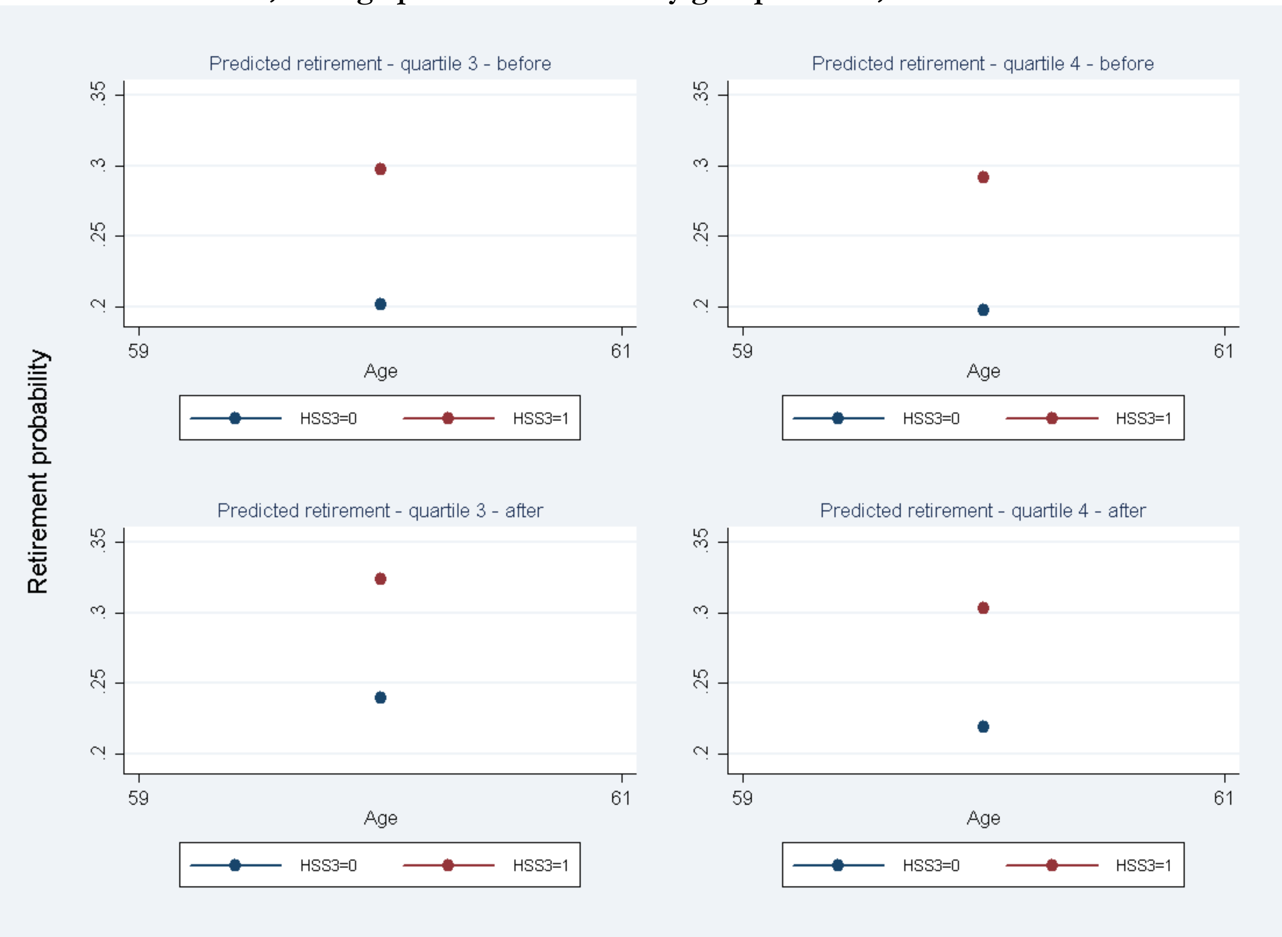
	HSP3				HSS3			
	1. quartile	2. quartile	3. quartile	4. quartile	1. quartile	2. quartile	3. quartile	4. quartile
	Model 19	Model 20	Model 21	Model 22	Model 23	Model 24	Model 25	Model 26
	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.	margins/Std.
OV	-2.3977***	-3.0106***	-3.1181***	-3.7332***	-2.3970***	-3.0276***	-3.1084***	-3.7115***
	(0.0524)	(0.0720)	(0.1029)	(0.1288)	(0.0519)	(0.0721)	(0.1025)	(0.1287)
HSP3	-0.0047	0.0022	-0.0015	-0.0034				
	(0.0025)	(0.0033)	(0.0038)	(0.0046)				
HSS3					-0.0018	-0.0026	-0.0040	0.0049
					(0.0027)	(0.0040)	(0.0036)	(0.0045)
HSP3 * OV	0.0122***	0.0073*	0.0081*	0.0130**				
	(0.0021)	(0.0032)	(0.0037)	(0.0048)				
HSS3 * OV					0.0126***	0.0139***	0.0148***	0.0069
					(0.0023)	(0.0039)	(0.0034)	(0.0047)
All controls	x	x	x	x	x	x	x	x
Observations	29,701	29,921	29,912	29,924	29,701	29,921	29,912	29,910
Pseudo R2	0.242	0.237	0.219	0.250	0.243	0.238	0.220	0.250

Figure 3b: Predicted retirement probability split by life annuity group, health shock, and before and after the reform, average person in life annuity group 1 and 2, HSS



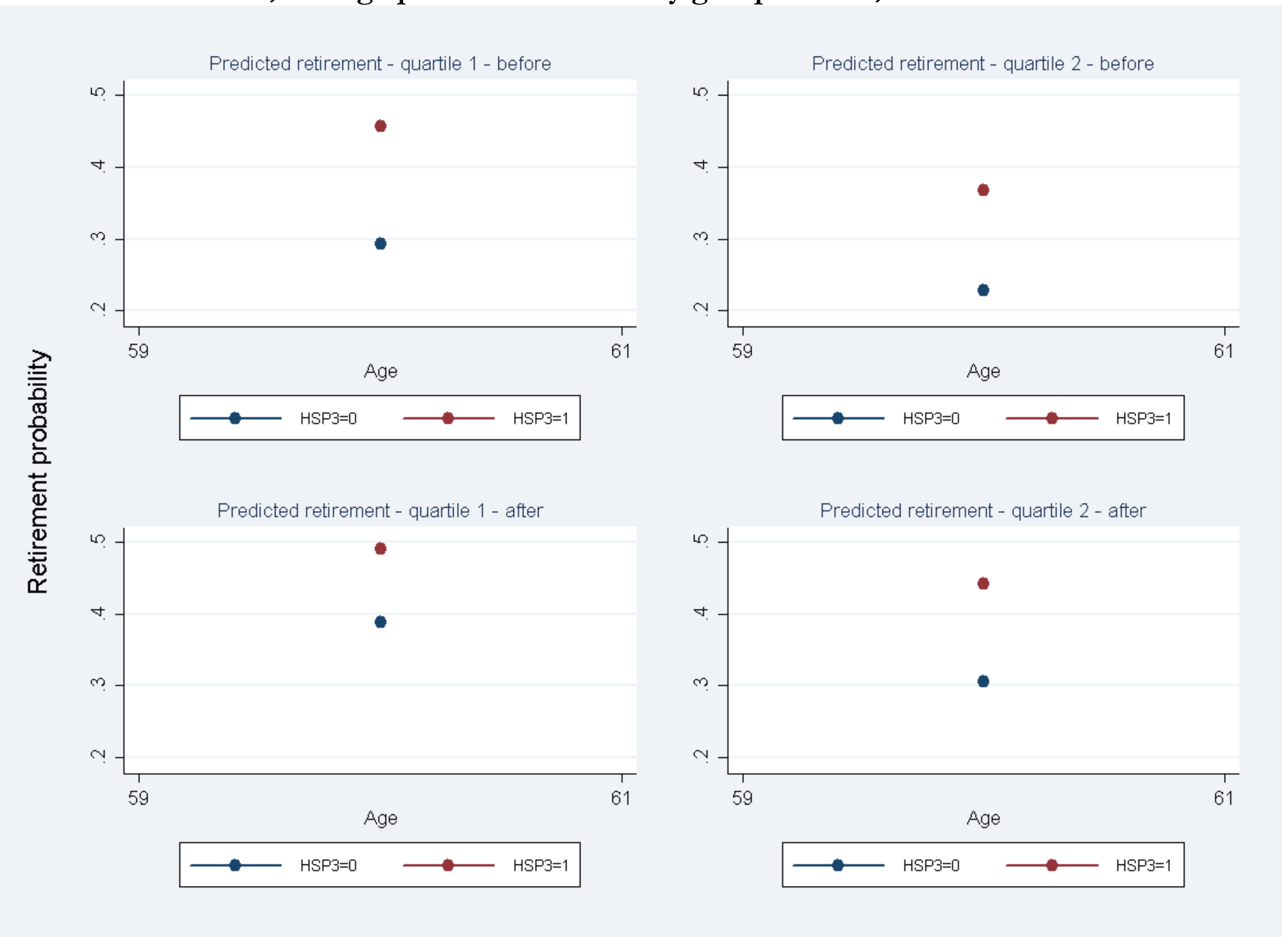
Note: HSS3 is a health shock to self according to definition 3 (defined by reduction in survival probability 5 years out from a new diagnosis). HSP3 is a health shock to partner according to definition 3.

Figure 4b: Predicted retirement probability split by life annuity group, health shock, and before and after the reform, average person in life annuity group 3 and 4, HSS



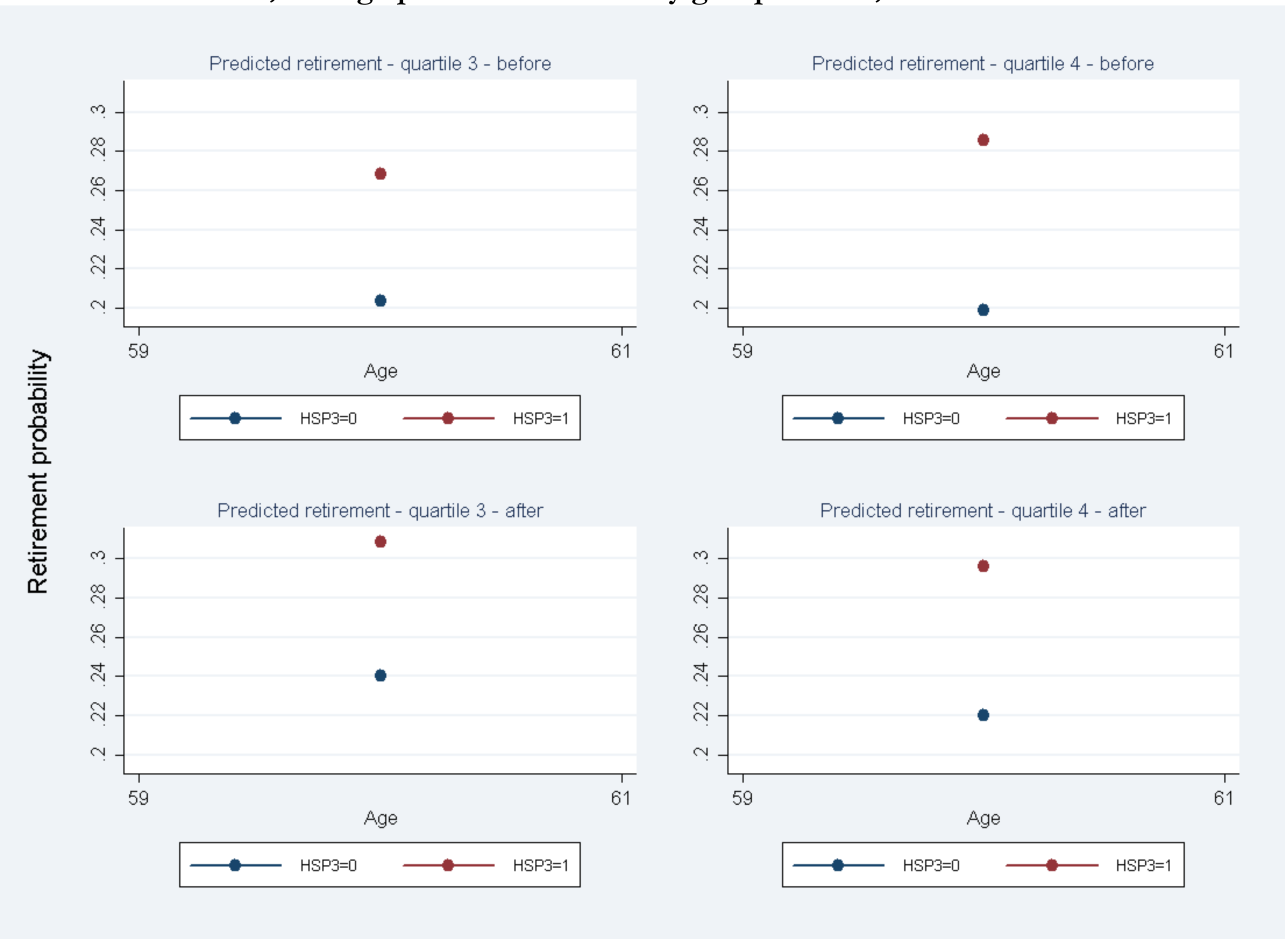
Note: HSS3 is a health shock to self according to definition 3 (defined by reduction in survival probability 5 years out from a new diagnosis). HSP3 is a health shock to partner according to definition 3.

Figure 5b: Predicted retirement probability split by life annuity group, health shock, and before and after the reform, average person in life annuity group 1 and 2, HSP



Note: HSS3 is a health shock to self according to definition 3 (defined by reduction in survival probability 5 years out from a new diagnosis). HSP3 is a health shock to partner according to definition 3.

Figure 6b: Predicted retirement probability split by life annuity group, health shock, and before and after the reform, average person in life annuity group 1 and 2, HSP



Note: HSS3 is a health shock to self according to definition 3 (defined by reduction in survival probability 5 years out from a new diagnosis). HSP3 is a health shock to partner according to definition 3.

Table 4b: Observed retirement probability split by life annuity group, before and after reform and health shock (HSS3)

Life annuity quantile	Minimum quantile kr.	Maximum quantile kr.	Reform	Health shock	Observations	Frequency health shock %	Average life annuity payout kr.	Retirement probability
1	0	14,506			14,119	2.9	7,897	41.5
2	14,508	32,928			13,723	3.5	22,578	36.0
3	32,936	80,528			13,705	3.2	52,173	28.7
4	80,544	472,000			13,959	3.0	168,011	24.6
1	0	14,506	Before	No	4,361		7,054	39.8
			Before	Yes	140	3.1	6,717	46.4
			After	No	9,348		8,290	42.2
			After	Yes	270	2.8	8,530	44.4
2	14,508	32,928	Before	No	2,752		22,360	35.7
			Before	Yes	111	3.9	22,892	39.6
			After	No	10,487		22,660	35.9
			After	Yes	373	3.4	21,807	38.3
3	32,936	80,528	Before	No	2,602		52,142	31.7
			Before	Yes	82	3.1	51,314	40.2
			After	No	10,669		52,196	27.9
			After	Yes	352	3.2	51,881	30.1
4	80,544	472,000	Before	No	2,848		169,838	30.7
			Before	Yes	101	3.7	166,307	33.7
			After	No	10,687		167,467	22.9
			After	Yes	323	2.9	170,408	24.5

Note: HSS3 is a health shock to self according to definition 3 (defined by reduction in survival probability 5 years out from a new diagnosis.

Table 5b: Predicted retirement probability split by life annuity group, before and after the reform, and health shock, average person in life annuity group

Life annuity quantile	Minimum quantile kr.	Maximum quantile kr.	Reform	Health shock	Observa-tions	Frequency health shock %	Retirement probability	
							HSS3	HSP3
1	0	14,506			14,119	2.9	0.36	0.36
2	14,508	32,928			13,723	3.5	0.29	0.29
3	32,936	80,528			13,705	3.2	0.23	0.24
4	80,544	472,000			13,959	3.0	0.22	0.22
1	0	14,506	Before	No	4,361		0.29	0.29
			Before	Yes	140	3.1	0.47	0.46
			After	No	9,348		0.39	0.39
			After	Yes	270	2.8	0.51	0.49
2	14,508	32,928	Before	No	2,752		0.23	0.23
			Before	Yes	111	3.9	0.34	0.37
			After	No	10,487		0.31	0.31
			After	Yes	373	3.4	0.41	0.44
3	32,936	80,528	Before	No	2,602		0.20	0.20
			Before	Yes	82	3.1	0.30	0.27
			After	No	10,669		0.24	0.24
			After	Yes	352	3.2	0.32	0.31
4	80,544	472,000	Before	No	2,848		0.20	0.20
			Before	Yes	101	3.4	0.29	0.29
			After	No	10,687		0.22	0.22
			After	Yes	323	2.9	0.30	0.30

Note: HSS3 is a health shock to self according to definition 3 (defined by reduction in survival probability 5 years out from a new diagnosis). HSP3 is a health shock to partner according to definition 3.

Appendix C: Illustration of the changed means testing due to the PEW reform in 1999

Table 1c show how different levels and types of pension were means tested in the PEW benefits after the reform in 1999. The means testing before the reform is equivalent to the rows "Equal to or older" than age 62, because only life annuity payouts were means tested (when paid out) before the reform.

Table 1c: Examples of means testing in the PEW after the reform in 1999

Age 62	**Pension scheme**	**Is the pension payed out**	**Pension type**	**Means testing in PEW, Euro**						
Pension wealth euro				**50,000**	**100,000**	**200,000**	**300,000**	**400,000**	**500,000**	**600,000**
Less than	Occupational pension	No	Capital	482	2,013	5,074	8,135	11,197	14,258	17,319
		No	Fixed annuity	482	2,013	5,074	8,135	11,197	14,258	17,319
		No	Life annuity	482	2,013	5,074	8,135	11,197	14,258	17,319
		Yes	Capital	482	2,013	5,074	8,135	11,197	14,258	17,319
		Yes	Fixed annuity	1,594	3,189	6,378	9,566	12,755	15,944	19,133
		Yes	Life annuity	1,594	3,189	6,378	9,566	12,755	15,944	19,133
	Private	No	Capital	482	2,013	5,074	8,135	11,197	14,258	17,319
		No	Fixed annuity	482	2,013	5,074	8,135	11,197	14,258	17,319
		No	Life annuity	482	2,013	5,074	8,135	11,197	14,258	17,319
		Yes	Capital	482	2,013	5,074	8,135	11,197	14,258	17,319
		Yes	Fixed annuity	482	2,013	5,074	8,135	11,197	14,258	17,319
		Yes	Life annuity	482	2,013	5,074	8,135	11,197	14,258	17,319
Egual to or older	Occupational pension	No	Capital	0	0	0	0	0	0	0
		No	Fixed annuity	0	0	0	0	0	0	0
		No	Life annuity	0	0	0	0	0	0	0
		Yes	Capital	0	0	0	0	0	0	0
		Yes	Fixed annuity	1,754	3,508	7,015	10,523	14,031	17,538	21,046
		Yes	Life annuity	1,754	3,508	7,015	10,523	14,031	17,538	21,046
	Private	No	Capital	0	0	0	0	0	0	0
		No	Fixed annuity	0	0	0	0	0	0	0
		No	Life annuity	0	0	0	0	0	0	0
		Yes	Capital	0	0	0	0	0	0	0
		Yes	Fixed annuity	0	0	0	0	0	0	0
		Yes	Life annuity	0	0	0	0	0	0	0

References

Banks J, Emmerson C, and Tetlow G (2007): Healthy Retirement or Unhealthy Inactivity: How Important Are Financial Incentives in Explaining Retirement? Working Paper, JEL Classification J26.

Barton M, Pissarides C, and Zabalaza A (1980): Social Security and the Choice between Full-Time Work, Part-Time Work, and Retirement, Journal of Public Economics, Vol. 14, pp. 245-276.

Belloni M (2008): The Option Value model in the Retirement Literature: The Trade-Off Between Computational Complexity and Predictive Validity. ENEPRI, Empirical research report No. 50.

Belloni M and Alessie R (2009): The Importance of Financial Incentives on Retirement Choices: New Evidence for Italy. Labour Economics, Vol. 16, Issue 5, pp. 578-588.

Belloni M and Alessie R (2012): Retirement in Italy: What an Option Value Model Tells Us. Oxford Bulletin of Economics and Statistics, Vol. 74, Issue 4.

Bingley P, Datta Gupta N, and Pedersen PJ (2004): The Impact of Incentives on Retirement in Denmark. In: Gruber J and Wise D (Eds.): Social Security Programs and Retirement around the World: Micro-Estimation, University of Chicago Press: Chicago, IL, pp. 153-234.

Bingley P and Lanot G (2007): Public Pension Programmes and the Retirement of Married Couples in Denmark. Journal of Public Economics. Vol. 91, pp. 1878-1901.

Bound J (1991): Self-Reported versus Objective Measures of Health in Retirement Models. Journal of Human Resources, Vol. 26, Issue 1, pp. 106-138.

Bound J and Burkhauser RV (1999): Economic Analysis of Transfer Programs Targeted on People with Disabilities. In: Ashenfelter O and Card D (Eds.): Handbook of Labor Economics, Vol. 3C, North-Holland: Amsterdam, pp. 3417-3528.

Burkhauser RV, Butler JS, and Gumus G (2003): Option Value and Dynamic Programming Model Estimates of Social Security Disability Insurance Application Timing. IZA Discussion Paper No. 941, November 2003.

Butrica BA, Johnson RW, and Mermin GBT (2009): Do Health Problems Reduce Consumption at Older Ages? CRR at Boston College. Working Paper 2009-9.

Börsch-Supan A, Brugiavini A, Jürges H, Mackenbach J, and Weber G (2005): Health, Ageing and Retirement in Europe. Manheim Research Institute for the Economics of Aging.

Casanova M (2010): Happy Together: A Structural Model of Couples' Joint Retirement Choices. Working paper, UCLA.

Christensen BJ and Lamb MK (2010): The Impact of Health Changes on Labor Supply: Evidence from Merged Data on Individual Objective Medical Diagnosis Codes and Early Retirement Behavior, CREATES Research paper 2010-62, pp. 1-48.

Coile CC (2004a): Health Shocks and Couples' Labor Supply Decisions. NBER, Working Paper 10810.

Coile CC (2004b): Retirement Incentives and Couples Retirement Decisions. Topics in Economic Analysis and Policy, The Berkeley Electronic Press, Vol. 4, Issue 1, Article 17.

Conley D and Thompson JA (2011): Health Shocks, Insurance Status and Net Worth: Intra- and Inter-Generational Effects. NBER, Working Paper 16857.

Datta Gupta N, Kleinjans KJ, and Larsen M (2012): The Effect of an Acute Health Shock on Work Behavior: Evidence from Different Health Care Regimes, JEL codes I12, I18, J26.

Deschryvere M (2005): Health and Retirement Decisions an Update of the Literature. ENEPRI, European Network of Economic Policy Research Institutes. Report No. 6.

Disney R, Emmerson C, and Wakefield M (2003): Ill Health and Retirement in Britain: A Panel Data Based Analysis. The Institute for Fiscal Studies. Working Paper 03/02.

Dowd B, Norton E and Karaca-Mandic P (2011): Interaction Terms in Nonlinear Models. Health Services Research. Vol. 47, Issue 1 pt. 1, pp. 255-274.

Dwyer DS and Hu J (2000): Retirement Expectations and Realizations: The Role of the Health Shocks and Economic Factors. In: Mitchell OS, Hammond PB, Rappaport AM (Eds.): Forecasting Retirement Needs and Retirement Wealth, University of Pennsylvania Press: Philadelphia, PA, pp. 274-287.

Groves R and Couper M (1998): Nonresponse in Household Interview Surveys. New York. Wieley.

Gruber J and Wise D (2004): Social Security Programs and Retirement around the World: Micro-Estimation, The University of Chicago Press.

Gustman AL and Steinmeier TL (2000): Retirement in Dual-Career Families: A Structural Model. Journal of Labor Economics, Vol. 18, No. 3, pp. 503-545.

Gustman AL and Steinmeier TL (2004): Social Security, Pensions and Retirement Behaviour within the Family. Journal of Applied Econometrics, Vol.19, Issue 6, pp. 723-737.

Gustman AL and Steinmeier TL (2005): The Social Security Early Entitlement Age in a Structural Model of Retirement and Wealth. Journal of Public Economics, Vol. 89, Issues 2-3, pp. 441-463.

Hagan R, Jones AM, and Rice N (2009): Health and Retirement in Europe. International Journal of Environmental Research and Public Health, Vol. 6, Issue 10, pp. 2676-2695.

Lazear E (1986): Retirement from the Labor Force: In Ashenfelter O and Layard R (Eds.): Handbook of Labour Economics, Vol. 1, Elsevier Science Publishers, BV, Amsterdam, pp. 305-355.

Lindeboom M and Van Doorslaer E (2004): Cut-Point Shift and Index Shift in Self-Reported Health. Journal of Health Economics, Vol. 23, Issue 8, pp.1083-1099.

Lumsdaine RL, Stock JH, and Wise DA (1990): Three Models of Retirement: Computational Complexity versus Predictive Validity. Topics in the Economics of Aging, NBER Project Report.

McClellan M (1998): Health Events, Health Insurance and Labor Supply: Evidence from the Health and Retirement Survey. In: Wise DA (Ed.): Frontiers in the Economics of Aging, University of Chicago Press: Chicago. IL, pp. 301-350.

McGarry K (2004): Health and Retirement: Do Changes in Health Affect Retirement Expectations? The Journal of Human Resources, Vol. 39, No. 3, pp. 624-648.

McGeary KA (2009): How do Health Shocks Influence Retirement Decisions? Review of Economics of the Household, Vol. 7, Issue 3, pp. 307-321.

Samwick A and Wise D (2003): Option Value Estimation with Health and Retirement Study Data. In: Ogura S, Tachibanaki T, and Wise D (Eds.): Labor Markets and Firm Benefit Policies in Japan and the United States, Chicago Press, pp. 205-208.

Stock J and Wise D (1990a): Pensions, the Option Value of Work and Retirement, Econometrica, Vol. 58, No. 5, pp. 1151-1180.

Stock J and Wise D (1990b): The Pension Inducement to Retire: An Option Value Analysis. In: Wise D (Ed.): Issues in the Economics of Aging. Chicago: University of Chicago Press.

Can We Reform Disability Insurance and Increase Labor Supply?

Paul Bingley
Michael Jørgensen

Abstract

Disability Insurance (DI) benefits may distort labor supply decisions for those capable of working because defining and observing the true health of an applicant is difficult. This difficulty might encourage moral hazard on the applicant's part. Moral hazard problems may be reduced directly through stricter screening of applicants or by reducing over-all DI benefits. We investigate the effects of attempts to mitigate moral hazard in the DI program by raising the stringency of the screening process for applicants. We measure the effect on DI awards and labor supply of two DI reforms that took place in Denmark during the 1990s: (1) state audits of local awards, (2) reduction in state rebates to local authorities for DI awards to applicants aged 60-66. Benefit generosity at the individual level remained unchanged throughout the estimation period, keeping economic incentive effects stable. State audits of local county boards showed that almost no counties awarded too few DI pensions according to the state audit board, whereas many awarded too many, and when these high awarding counties were audited, they increased rejection rates the following year by on average 21%. Reduced state rebates to municipalities for DI awards for those aged 60-66 increased rejection rates by 5 percentage points for this age group compared to those aged < 60. Instrumental variables estimates show that a 10% rise in rejection rates leads to a 6.4% increase in labor supply. We find evidence that those awarded DI experienced a lower increase in survival probability, indicating that the increased stringency managed to reject the healthier applicants.

1. Introduction

In Western Europe and North America men's labor supply has been decreasing since the 1960s, among other reasons because a bigger percentage of the potential working population is receiving public transfers—especially disability benefits. The same development has taken place in Denmark.[48] This almost continuous increase in DI recipients over the last 3-4 decades, despite that general health and life expectancy has been improving (see e.g. Autor and Duggan, 2003; Autor and Duggan, 2006), has made researchers (see e.g. Parson, 1980; Bound, 1989; Gruber and Kubik, 1997; Bound and Burkhauser, 1999; Karlström et al., 2008; De Jong et al., 2011) argue that generous benefits, combined with the difficulty that awarding authorities have in defining and observing disability, creates moral hazard[49] problems that distort labor supply.

The two broad policy approaches to tackling the moral hazard problem have been improved targeting of awards through increased stringency and a reduction in benefits, so that work becomes relatively more attractive for all but the "truly disabled". While most of the literature on disability reforms has focused on the labor supply effect from a reduction in benefits (se e.g. Bound and Burkhauser, 1999; Gruber, 2000; Chen and Van der Klaauw, 2008), we focus on two stringency reforms of DI. We focus on stringency reforms, because earlier studies (se e.g. Bound, 1989; Gruber and Kubik, 1997; Chen and Van der Klaauw, 2008) have argued that DI replacement rates are correlated with past earnings, which makes it difficult to separate the effect of a weak labor force attachment from the effect of generosity of DI benefits on labor supply.

Researches share the inability of the awarding authority to observe true individual DI applicant health and disability status, which creates variation in award rates across awarding authorities (municipalities). Besides the variation in award rates among municipalities due to unobserved variation, municipalities also engage in different awarding practices for a number other reasons–expertise in the awarding process, resources, number of cases due to demographics etc.

Because of the large variation in municipalities DI award rates and the general increase in DI awards, a number of DI reforms were initiated in the 1990's to make the screening process more streamlined across municipalities and to reduce the number of awards and increase labor supply. We analyze two of

[48] From 1971 to 1995 the percentage of people in Denmark aged 16-66 receiving disability benefits doubled from 4% to 8%.
[49] Throughout the paper we will use the term moral hazard to describe the situation in which individuals are awarded DI because the awarders can not observe the applicants true health (asymmetric information). This phenomenon might also be called adverse selection. We will call it moral hazard since that's the term that is used in most of the literature.

these reforms to the administration of the awarding process: a) a 1995 audit reform and b) a 1997 rebate reform. The 1995 audit reform aimed at disseminating best practices by auditing a random sample of award decisions for a rolling 2 out of 15 counties each year through 2002. The 1997 rebate reform incentivized municipalities to increase screening stringency by reducing state rebates for municipal DI benefit payments. Before the 1997 reform the state rebate was 50% for those aged 16-59, but for those aged 60 and above it was 100%. Post-reform state rebates were 50% for all ages. For individuals receiving DI benefits in the 1990s, no changes in benefit generosity occurred, thereby making our period of analysis especially informative about administrative targeting effects.

Our paper combines the insight of Gruber and Kubik (1997) on changes in DI application rejection rates and labor supply with the insight of Bound (1989) about the labor supply of individuals who had applications for DI rejected. We investigate the effects of attempts to mitigate moral hazard in the DI program by raising the stringency of the screening process for applicants. We look at reforms at the interface between municipalities and the state government (the 1995-2001 audit reform and the 1997 rebate reform). Audits were directly targeted at specific counties and municipalities. Rebate and screening reforms affected municipalities differently according to the pre-reform age-balance of awards and screening of applicants for remaining work capacity, respectively. Difference-in-difference (DID) analyses at the municipal-year level show that both reforms reduced municipality DI award rates. We use the reforms to provide exogenous variation in award probability in an instrumental variables analysis of labor supply, in order to estimate a local average treatment effect of DI denial for those whose award could be changed by the award stringency changes.

We contribute to the existing literature in several ways: First we analyze the effect of two different stringency reforms in a DID setup, which allows us to evaluate two different forms of stringency tools– auditing (best practice) vs. reduced rebates (economic incentive). This comparison is valuable to policymakers who have to decide how to moderate DI awards. Secondly we combine the partial analyzes of earlier studies in one study: a) we use two natural experimental stringency reforms to generate exogenous variation in rejection rates and characterize their effect on rejection rates in a DID setup, b) we estimate the labor supply effect of changing rejection rates, due to reforms of DI in the 1990's, in a two-stage least square regression design, and c) we evaluate if increased stringency primarily excluded the most healthy DI applicants.

We find that auditing increased rejection rates by 21% on average for those municipalities that awarded too many DI's according to best practice, and lower rebates for the age group 60-66 increased rejection rates by 4.6 percentage points for this age group relative to those aged 50-59. Our two-stage instrumental variable estimates of labor supply show that a 100% increase in instrumented rejection rates increase labor supply by 64%. And finally our analysis shows that greater stringency resulted in increased rejection of the healthiest applicants.

The remainder of the paper is organized as follows. The next section presents a literature review on DI and labor supply. Section 2 presents the institutional background, disability programs, and reforms. Section 3 describes our data sets, section 4 presents descriptive statistics. Section 5 presents our estimation design and results, and section 6 concludes.

1.1 Literature review

To obtain a causal estimate of the disincentive effect of DI on work and earnings, an estimate of the counterfactual work and earnings of DI recipients is needed. Several approaches have been followed which can be grouped into three general identification arguments. A first group of studies, following the work of Bound (1989), simply compares awarded and denied DI applicants. A second group, following on from Gruber and Kubik (1997) uses natural experiments to provide groups according to region, examiner, or age group differences in award propensity and conducts reduced form difference-in-differences and instrumental variables analyses. A third approach uses experimental variation in award propensity to conduct reduced-form difference-in-differences analysis.

Bound (1989) was the first to compare DI recipients with denials in order to estimate an upper bound on the counterfactual labor supply of recipients. This was an upper bound because, despite being otherwise observationally quite similar, denials obviously on average suffer from slightly fewer severe impairments than awards. Parsons (1991) criticized Bound (1989) on the grounds that the two groups are observationally different, that a significant minority of initial denials are subsequently awarded on appeal, and that applying itself diminishes work capacity. Bound (1991) replied to re-iterate the claim originally stated. However, Parson's points regarding the ins-and-outs of the awarding process resonated and have motivated very recent work.

Chen and van der Klaauw (2008) repeat Bound (1989) for more recent years and focus on a marginal group of applicants whose claim is not decided on medical grounds alone in a first round, but where vocational factors are also considered. The argument here is that this marginal group, though less representative of the general pool of applicants, is more relevant for policy and provides a tighter upper bound on the potential labor supply of awards. Wachter et al. (2011) repeat the Bound (1989) analysis on a larger and more comprehensive administrative dataset for more years.

Gruber and Kubik (1997) use variation in state-level reductions in award probabilities to estimate the reduced form difference-in-differences relationship between award probability and labor supply.

Masteas et al. (2011) compare subsequent work of DI applicants who were initially accepted or denied awards because their cases were dealt with by disability examiners with different propensities to award. They use instrumental variables to identify a LATE for applicants whose status could be changed by examiner allocation.

French and Song (2011) follow a similar strategy to Masteas et al. (2011) but use award propensities of administrative law judges in the decision appeal process as instruments for eventual DI receipt in a labor supply equation. This is similar to the second part of our study, but French and Song do not characterize the effect of the natural experiment in a DID setup.

Autor et al. (2011) use the complement of the French and Song (2011) data, instead containing those initially awarded DI benefits to analyze the effect of decision processing time on subsequent labor supply. Average examiner decision time is used to instrument individual award times.

Borghans et al. (2012) analyze substitution between social assistance programs in response to re-examination of DI recipients below an age threshold in the Netherlands. In a reduced form difference-in-difference framework, they find that individuals react to DI generosity and stringency changes by replacing income with earnings or other transfers.

Similarly, Karlström et al. (2008) analyze removal of special DI eligibility rules over an age threshold in Sweden. Using a difference-in-differences framework they find substitution to alternative programs for sickness and unemployment insurance.

De Jong et al. (2011) analyze an experiment in stricter screening of DI applications in some Dutch regions. In a reduced form difference-in-differences analysis they find fewer DI applications and more return to work after sickness absence, largely due to sanctions on employers lacking re-integration efforts for long-term sick workers. They also find indication of efficient targeting–stricter screening did not harm the unhealthiest.

In this paper we follow two complementary identification approaches. First age-specific rebates from state to local government were changed and this provides age-group specific stringency incentives along the lines of Karlström et.al (2008) and Borghans, et.al. (2012). This approach falls into the general class of group-specific stringency variation. We observe municipalities who are the relevant decision-making authorities (analogous to States in Gruber and Kubik (1997)), but we are unable to observe examiners or caseworkers on a consistent basis and so are unable to repeat Masteas, et.al (2011). Furthermore we have a regional award decision audit natural experiment along the lines of the experimental variation in Borghans, et.al. (2012).

Reduced form difference-in-differences is the most common empirical strategy of the second and third approaches reviewed, where labor supply outcomes are regressed on policy-driven stringency variation. The exceptions are studies by Autor, et al. (2011), Mateas, et al. (2011) and French and Song (2011) which use examiner or judge award propensity to provide instruments for individual DI awards and which then identify a LATE from a second stage labor supply equation.

We follow a hybrid approach because of the different nature of our two sources of exogenous award probability variation. In a first set of DID analyses we use age-group specific stringency changes in instantaneous individual award probability. In a second set of difference-in-difference analyses we use region-specific award decision audits on individual award probability in the following years. In a final set of instrumental variables regressions we use municipality-year specific variation in award propensity to determine individual award probability in a first stage regression to identify a LATE from a second stage individual labor supply equation. Simulations allow us to illustrate how each stringency policy tool affects award probabilities and thereby differentially affects labor supply decisions.

Our contribution is to combine two different natural experiments that used two different forms of instruments to increase rejection rates–best practice through auditing and economic incentives through reduced rebates to municipalities. We first characterize the effects of these experiments in a DID design to demonstrate that they managed to increase rejection rates. Then we estimate a two-stage least square labor supply equation to demonstrate how an increase in instrumented individual rejection rates affected labor supply. And finally we evaluated if the stringency reforms managed to increase rejection for the healthiest applicants (targeting efficiency). The first part allows us to evaluate the effect of two different stringency tools, which has not been done before, and it allows policymakers to compare effects of different tools. The second part allows us to characterize by how much an increase in instrumented individual rejection rates increased labor supply. The third part will show us if increased stringency evaluated the healthiest applicants or just made the unhealthiest applicants worse off, which is important for targeting strategies and efficiency. It is the first time that two different types of natural experimental stringency variation have been measured in the same context. Inference should be informative about both comparison of policy tools and empirical methods.

2. Disability program, reforms and development

In this section we describe the DI program and relevant reforms of it during the 1990's. We also illustrate the development in the number of recipients, the number of applicants and the rejection rate over time.

2.1 Disability program and reforms

The basic structure of DI benefits in Denmark during our period of analysis (1992-2002) was put in place in 1984. DI was awarded at three levels, according to medical criteria for the highest levels of support, and on medical or social grounds for the lowest level. At that time, local government was organized into 15 counties, comprising 275 municipalities.

In the early 1990s the municipalities, together with the DI applicants, prepared a case for DI and presented it to a county board. The state government rebated all the municipality's DI expenditures. When controlling for caseload (demographic characteristics of applicants), Bengtsson (2002) found a wide range of application rejection rates between counties—7-32%. This variation in award stringency motivated a pilot study where a few municipalities themselves made award decisions. The pilot study led to the 1992 decentralization of the award process to the municipalities, with rebates on awards for those under age 60 reduced from 100% to 50%. The frequency of DI awards continued to increase, thus motivating three staggered reforms during the rest of the decade.

The first reform, in 1995, aimed at harmonizing the award decisions across municipalities through an audit of a random sample of 200 applications from 2 counties out of 15 every year from 1995 to 2002. The audit reform can be seen as a natural experiment where the two counties that were randomly picked for auditing was the treatment group and the other 13 counties constituted a natural control group. The 15 counties that were audited were divided into 275 municipalities that each awarded DI pensions. State lawyers re-assessed around 100 decisions in each county each year, and subsequently reported how many of the rulings they agreed on, and they did not agree on.[50] These reports allow us to differentiate between the number of cases where municipalities rejected DI and the state lawyers awarded DI (Type 1 errors) and those cases where municipalities awarded DI and the state lawyers rejected (Type 2 errors).

---

[50] Municipalities were expected to change their practice accordingly to the state lawyers recommendations, but the outcome of the assessed cases would not be changed.

As only 2 out of 15 counties were audited each year, the remaining counties constitute a natural control group. The counties did not know when they would be selected for auditing, but over time they knew that their chances of being selected increased, because selection was done without replacement. Therefore we exclude the last years of auditing in our estimations.

The second reform in 1997 reduced the state DI rebates to the municipalities. Before 1997 the state government rebated all DI expenses for those aged 60-66 but only half for those aged 16-59. After the reform, the rebate was 50% for all. The younger group is a natural control because of the similar rebate throughout the period.

In 1998 a filtering reform was introduced, which changed the DI award process so that all other programs needed to be considered before a DI application could be processed. Previously, one department in a municipality would prepare an application for DI benefits, together with the applicant, and another department in that same municipality would make the DI award decision. After 1998 all other programs—including subsidized job training, rehabilitation and disabled wage subsidies–had to be considered first, as part of a work test involving consideration of medical records and physical and psychological tests, before a DI application could be filled out.

Disabled wage subsidies (fleksjob), which was introduced with the 1998 reform, is a regular job at reduced work capacity with a wage subsidy given to the employer according to the degree of reduced work capacity. Disabled wage subsidies are almost equal to full DI because both programs are permanent in duration (once granted, there is no reassessment of status) and because the award grantee has to have a permanently reduced earnings capacity. The main difference between the disabled wage subsidies and full DI is that people still work part- or full-time while receiving the disabled wage subsidies[51], whereas those receiving full DI seldom work. The employer receives a wage subsidy according to the capacity reduction.

There existed three levels of wage subsidy in the period analyzed: 33%, 50% and 67%, and the state government rebates for the disabled wage subsidies were 50% until 1997, 100% from 1997 to 2001 and 65% from 2002. Rehabilitation was given to a person with reduced earnings capability when other vocationally or job-oriented programs cannot help a person to become self-supported. Its goal is to

[51] Individuals awarded disabled wage subsidies will receive unemployment benefit and sickness benefit, when they lose their jobs or when they get sick from work.

create or maintain a person's attachment to the labor market. Rehabilitation can consist of different initiatives, such as a short or long educational course, rehabilitation in private or public firms or help becoming self-employed. Rehabilitation can "only" be granted up to 5 years, because it is a temporary program¬—but normally it will be shorter. About 40% become self-supported after a rehabilitation course, 30% receive cash benefit of sickness benefit, and only 5% are granted flexjobs or DI benefit (Danish Economic Council, 2003).

Unlike in the 1995 audit reform and the 1997 rebate reform, there existed no natural control group for the 1998 filtering reform. Instead, we use the variation in the change in rejection rates across municipalities that the reform created in our first stage labor equation. The variation in the change in rejection rates arises because municipalities did not have the same filtering practice before 1998.

In 2003 the DI program was reformed, so that a single level for new awards replaced the former three levels of benefits. During the 1990s no changes occurred to the structure, and benefits were simply adjusted annually for wage inflation. To focus on the consequences of administrative changes, rather than on benefit level incentives, our study runs from 1992 through 2002.

2.2 Development in DI recipients and related programs

The DI reforms described in the previous section was motivated by a continuous increase in DI recipients from the introduction of the program in 1956[52] and until the early 1990s. In this section we describe the developments in DI recipients from the reform in 1984 and until the end of our estimation period.

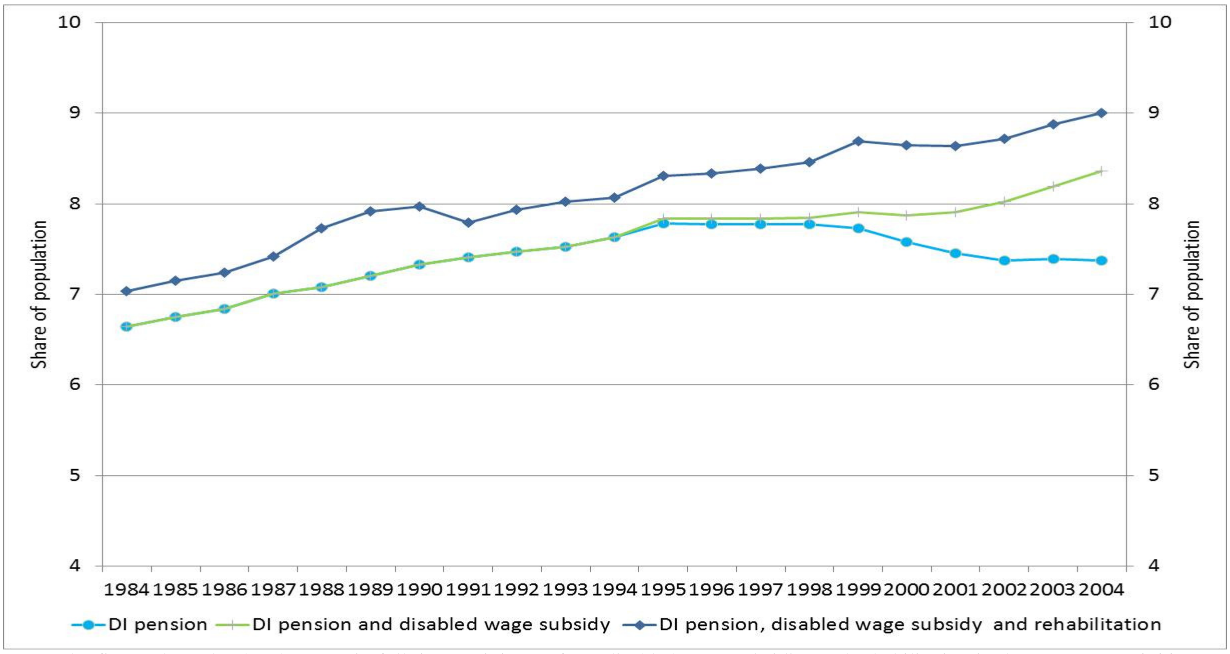
The description of the reforms in the previous section clarifies that it is important to consider the developments in disabled wage subsidies and rehabilitation together with DI after 1998.

Therefore, figure 1 presents the percentage of the population aged 16-66 receiving DI, disabled wage subsidies and rehabilitation benefits. The full DI recipient percentage rose from 6.6 in 1984 to 7.8 in 1995, falling back to 7.4 by 2004. Adding disabled wage subsidies (which were introduced in 1998) to the regular DI maintains the upward slope after 1998, and by 2004 8.4% at the population aged 16-6 received either DI or disabled wage subsidies. Even through the curve which includes the disabled wage subsidies maintains an upward trend, it is important to remember that those receiving disabled wage subsidies also worked and therefore supplied some amount of labor. So even if this group only

[52] The introduction in 1956, was a further development of an earlier program from 1933.

supplied 50% of labor, the upward trend we see from 1984 to 1995 would still be broken. Adding rehabilitation adds an almost parallel shift to the curve with disabled wage subsidies and full DI, and the total amount of full time individuals receiving one of the three benefits rose from 7% to 9% from 1984 to 2004.

Figure 1: Share of population aged 16-66 receiving DI, disabled wage subsidies and rehabilitation



Note: The figure show the development in full time recipients of DI, disabled wage subsidies and rehabilitation in the age group 16-66 as a share of the population aged 16-66.

The full DI recipient percentage rose from 6.6 in 1984 to 7.8 in 1995, falling back to 7.4 by 2004. Adding the disabled wage subsidies (which was introduced in 1998) to the regular DI maintains an upward slope after 1998, and by 2004 8.4% at the population aged 16-6 received either DI or disabled wage subsidies. Even through the curve including the disabled wage subsidies maintains an upward trend, it is important to remember that those receiving disabled wage subsidies also worked and therefore supplied some amount of labor. So even if this group only supplied 50% of labor, the upward trend we see from 1984 to 1995 would still be broken. Adding rehabilitation adds an almost parallel shift to the curve with disabled wage subsidies and full DI, and the total amount of full time individuals receiving one of the three benefits rose from 7% to 9% from 1984 to 2004.

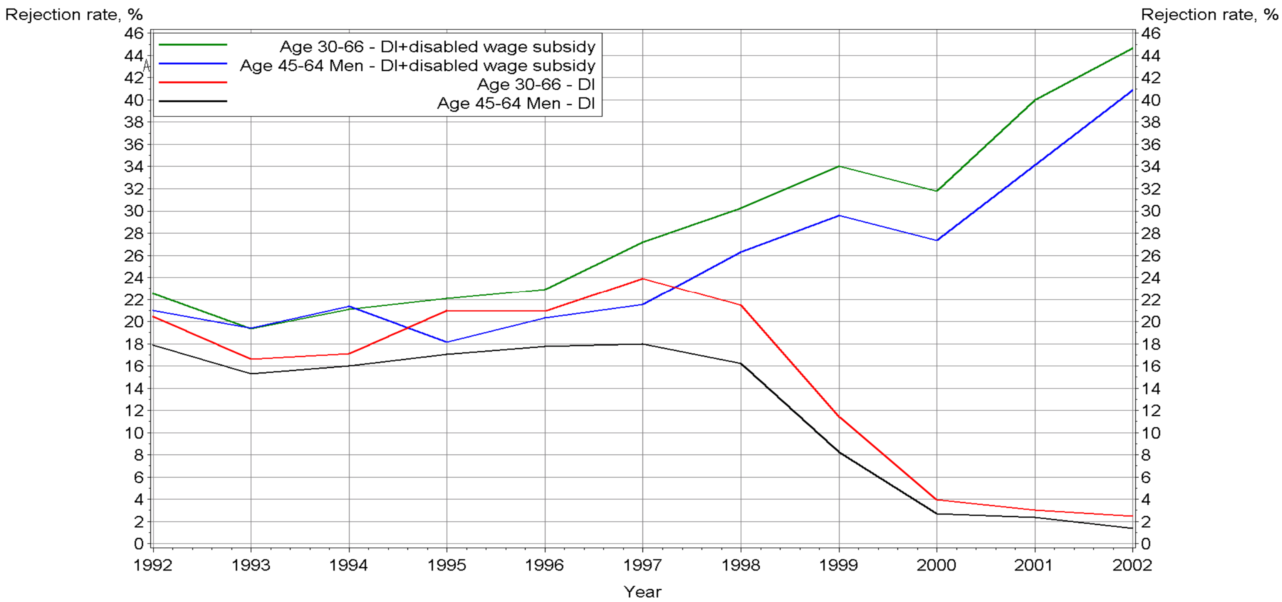
2.3 Development in applicants and rejection rates

The municipalities examine the DI applicant pool physically and mentally to establish the extent and expected duration of incapacity. From 1998 the government mandated that rehabilitation and disabled wage subsidies be considered first, and full DI be awarded as a last resort. Those assessed as having medium-term incapacity would join a rehabilitation program. The permanently disabled would be assessed for work capacity and awarded full or disabled wage subsidies accordingly. The remainder would be rejected for DI.

Our main analysis, throughout the paper, will be on the age group 30 to 66, but since Gruber and Kubiks (1997) estimated their model on 45-64 year old males, we also show the development in rejection rates for men aged 45 to 64 (their estimation sample). In addition, we calculate the rejection rate for DI applicants only and for DI applicants + disabled wage subsidies recipients, and consider those awarded disabled wage subsidies as rejected applicants for full DI, because they would most definitely have applied for full DI before the 1998 filtering reform. The main reason for including disabled wage subsidies recipients in the calculation of the rejection rate is to capture the effect of the 1998 filtering reform on rejection rates.

Figure 2 shows the development in the rejection rate from 1992 to 2002 for the age groups 30-66 and 45-64 (men) and for the two different definitions of the rejection rate. First we comment on the age group 30-66.

Figure 2: Rejection rate for DI and DI + disabled wage subsidies, different age groups

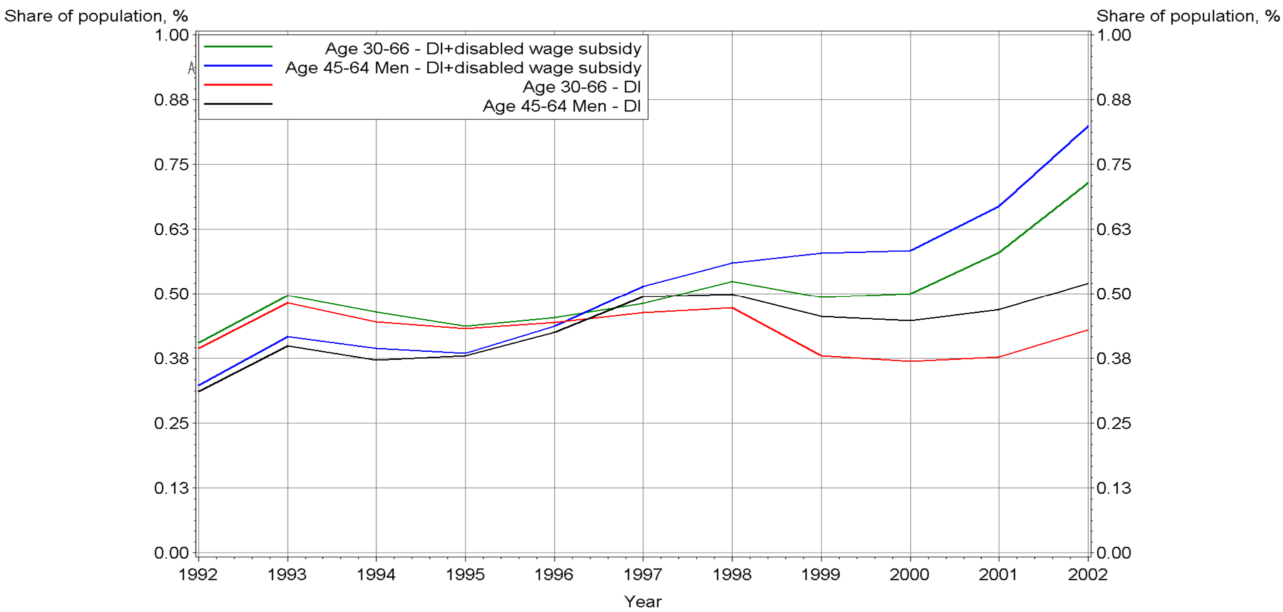


Note: The figure show the development in rejection rates calculated with and without disabled wage subsidies recipients as rejected applicants.

Because of the changed filtering of potential applicants after 1998, the rejection rate for DI fell drastically from 1998 (from 24% in 1997 to 2.5% in 2002), because a big share of those who would have been rejected prior to 1998 never applied for DI after the 1998 reform. If we look at the curve including disabled wage subsidies recipients as rejected full DI applicants, then the picture changes drastically. Now the rejection rate increases from 27% in 1997 to 45% in 2002.[53] We argue that because both full DI and disabled wage subsidies required a permanently reduced earnings capacity and because disabled wage subsidies was the last resort before full DI, the appropriate risk-set to calculate the rejection rate from is the one including disabled wage subsidies.[54]

If we look at the whole period from 1992 to 2002, then we see that the rejection rate (including disabled wage subsidies) increased steadily almost over the entire period, indicating that the three reforms managed to increase rejection rates. The development in rejection rates for the age group 45-64 (men) in figure 1 is very similar to the development for those aged 30-66. The only difference is that the rejection rate is a little bit lower for those aged 45-64 (for both definitions of the rejection rate).

Figure 3: Share of population applying for DI or DI + disabled wage subsidies, different age groups



---

[53] The reason why the curves with and without disabled wage subsidies does not overlap before 1998 is that there existed a program called 50/50 before 1998 that were similar to disabled wage subsidies, which provided 50% wage subsidies for those with earnings capacity reduced permanently by 50% or more. Participation in the 50/50 program was stable until the 1998 reform that changed it into disabled wage subsidies and required that it be considered before an application for full DI.

[54] Because both full DI and disabled wage subsidies required a work test involving consideration of medical records and physical and psychological tests and both programs required a permanently reduced earnings capacity, we also argue that the expansion of the risk-set because of this new program (disabled wage subsidies) was minor.

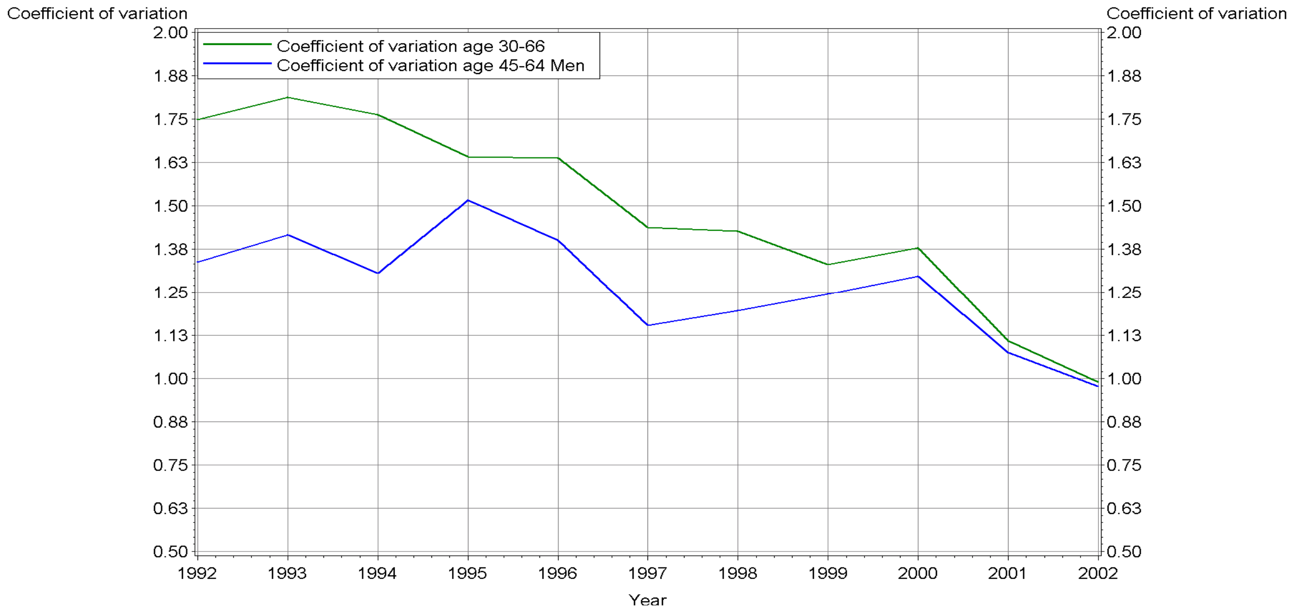
Figure 3 shows the development in applicants as a share of the population for the age groups 30-66 and 45-64 (men), for the two different definitions of the risk-set. The figure shows that the share of applicants for DI aged 30-66 was stable until 1998, when it then decreased, whereas there was an increase in the applicants for DI + disabled wage subsidies after 2000–indicating that the disabled wage subsidies attracted more applicants after 2000. The pattern is the same for those aged 45-64 (men) after 1998, but the share was already increasing for this group before the 1998.

2.4 Variation in rejection rates across municipalities over time

As previously discussed, the 1995 audit reform mandated that around 150 random cases from two counties each year be audited from 1995 to 2001. The goal was to help establish best practices in award decisions among municipalities, and to reduce the large variation in awarding rates that were found in the early 1990s. To measure the development of variation in the rejection rates between municipalities over time, we calculate the coefficient of variation (CV), which is defined as the standard error of the rejection rate between municipalities each year divided by the mean rejection rate in municipalities each year.

Figure 4 shows the development in the CV for the age groups 30-66 and 45-64 (men). If we first look at the curve for the age group 30-66, then we see that the coefficient of variation falls almost over the entire period from 1992 to 2002, indicating that the 1995 audit reform managed to harmonize the variation in rejection rates between municipalities.

Figure 4: Development in the coefficient of variation between municipality-year, rejection rates



The development in the curve for those aged 45-64 (men) is a bit different. For this age group the coefficient also decreases for most of the years and ends up at lower level in 2002, but there is an increase in the curve in both 1995 and 1998-2000. One possible explanation for the increase in 1998 might be the new variation in rejection rates that the disabled wage subsidies caused, because a relative bigger share of men received disabled wage subsidies.

3. Data

To create our estimation dataset we use three different sources of data. Our first source of data is population-based administrative longitudinal data containing information on demographics, income and health. Our second source (DREAM) holds information on weekly labor force attachment, and our third source is the data from the social security appeals board register of DI applications and awards.

The register data constitute a 50% random sample of the total Danish population and is available from 1990 and onwards. The registers holds information on background characteristics (age, gender, ethnicity, civil status, socioeconomic status etc.), economic measures (total income, income from work, total wealth, pension wealth etc.), health indicators (days in hospital, diagnoses, sickness absence from work).

DREAM is a longitudinal data set accounting for labor force attachment on a weekly basis since July 1 1991. DREAM includes almost every form of benefit a person can receive (cash benefits, sickness benefits, retirement benefits, unemployment benefits, disabled wage subsidies benefits, full DI benefits, labor market activation benefits, rehabilitation benefits, leave benefits etc.). We use these data to create a continuous labor force attachment measure.

The award and reject data set is from the state board receiving and treating complaints about administrative decisions. These data are available from 1992 onwards. The data set contains information on awards and rejections, whether it was the first application or an appeal, what level of DI that was applied for and different background information.

4. Descriptive statistics

In this section we show different descriptive statistics. We first compare the applicant sample to the population to see how they differ. Then we show the labor force attachment before and after applying for DI split by rejection status and finally we present the main results from the audit reports.

4.1 Descriptive statistic for applicants versus the population

When we compare the population characteristics to the application data we divide the application data into three groups: awarded full DI, awarded disabled wage subsidies and rejected for full DI. The group that received disabled wage subsidies is also considered as rejected for the full DI, but since the disabled wage subsidies program was part of the 1998 reform strategy to filter applicants before applying for DI, we show them separately.

Table 1 shows characteristics for the population, those awarded full DI, those awarded a disabled wage subsidies and those rejected for full DI, for the age group 30-66 from 1992-2002.

Table 1: Descriptive statistics, 1992-2002, age 30-66, All, Full DI, disabled wage subsidies and rejected

	All	Awarded Full DI	Awarded Disabled wage subsidies	Rejected
Observations	15,172,762	48,815	12,472	8,654
Age	45.3	51.8	45.3	47.5
Men	50.0	37.2	42.2	36.9
Single	15.9	29.3	23.4	28.6
Origin - Denmark	93.4	88.6	94.1	84.0
Lag(t-1) employment percentage	76.0	41.6	41.2	40.5
Lag(t-2) employment percentage	76.5	48.0	46.7	45.2
Lag(t-1) total income	340,253	195,034	255,417	171,202
Education level				
Primary school	29.2	57.3	42.1	58.4
Occupational/grammar school	42.8	28.9	41.4	26.8
Short highed education	5.1	2.1	3.4	2.0
Intermediate higher education	14.6	5.4	9.5	5.5
Long higher education	6.1	1.4	1.5	1.3
No diagnoses	90.4	76.8	84.0	81.4
Lag(t-1) sick percentage	1.5	25.2	31.2	17.2
Lag(t-1) days hospital sick	0.5	4.0	2.3	1.8

Note: The column All refers to the population. Full DI refers to those awarded DI. Disabled wage subsidies refers to those awarded disabled wage subsidies (rejected for full DI). Rejected refers to those who were rejected for full DI, and were not awarded disabled wage subsidies. The calculating of descriptive statistics for the population is based on longitudinal data, while the calculation for the groups Full DI, disabled wage subsidies and Rejected is based on the year the individual applied for DI–therefore there are much fewer observations in these columns.

Those awarded full DI are on average older than the population (52 vs. 45), while those awarded a disabled wage subsidies is the youngest group of applicants (45.3), and those rejected are also somewhat older than the general population. The share of men is lower in all three application groups compared to the general population. The share of singles in the three application groups is much higher than in the general population (almost double). The share of individuals originating from Denmark is lower in the group awarded full DI and in the rejected group, while the share is a little bit higher in the group awarded disabled wage subsidies than in the population.

The share of employment in year t-1 and t-2 is, as expected, much higher (almost twice as high) in the population compared to the three application groups, and the level only varies a little between the three application groups–the lowest share is in the group rejected. The share of employment goes down from t-2 to t-1 in all three application groups, indicating that they are becoming more separated from the labor market prior to applying for DI. Income in the previous year is much higher in the population than in the three application groups. The lowest income is in the rejected group (171,202 DKK) while it is highest in the population (340,253 DKK).

When we compare the distribution over educational groups, we see a very clear pattern: the two lowest educational groups constitute a much larger share in all three application groups than in the population as whole.

The last part of table 1 compares health measures. The share without a diagnosis is highest in the population (90.4%) and lowest in the group awarded full DI (76.8%). The share would be higher if psychological diagnoses were included, especially because the share receiving DI because of psychological diagnoses has been increasing over time.[55] Absence from work because of sickness[56] in t-1 is very different between the population and the three application groups. The average absence was 1.5% for the population while it was between 17.2% (rejected) and 31.2% (awarded disabled wage subsidies) in the three application groups. Those awarded DI spent on average 4 days in hospital in t-1 while the population average was 0.5 days–the group awarded disabled wage subsidies and the group rejected spent on average 2.3 and 1.8 days in hospital. The three health measures show that those awarded full DI had more diagnoses and spent more days in hospital, while those awarded disabled wage subsidies had more sick days.

---

[55] It is also important to remember that an individual can be granted DI because of social circumstances.
[56] Absebce is only measured beyond 14 days, because no refund from state is given before then.

To sum up: From table 1 we see that there are more women, more singles, more people with lower education levels, more people with low incomes in t-1 and t-2, more people with a diagnosis, more people that were sick in t-1, and more people who spent days in hospital in t-1 in the application group (awarded full DI, awarded disabled wage subsidies and rejected) than in the population. This indicates that applicants were motivated to apply for DI for both health and economic reasons (moral hazard).

4.2 Labor market status after applying

More studies (see e.g. Parson, 1991; Benitez-Silva et al., 1999) have argued that a share of those rejected for full DI would appeal their cases and be awarded DI later. We therefore show the labor market attachment in t+1, t+3 and t+5 after applying, for those awarded full DI, awarded a disabled wage subsidies and those rejected. Table 2 shows the share of time employed, the share of time receiving disabled wage subsidies and the share of time receiving full DI in t+1, t+3 and t+5 after applying for full DI, for the age group 30-66 in the period 1992-2002.

Table 2: Labor supply in t+1, t+3 and t+5 after applying for full DI, 1992-2002, age 30-66, Divided by award decision

	Awarded Full DI	Awarded Disabled wage subsidies	Rejected
Observations	48,815	12,472	8,654
Lead(t+1) employment	2.4	77.8	34.1
Lead(t+3) employment	2.4	70.6	27.9
Lead(t+5) employment	2.7	66.7	23.8
Lead(t+3) disabled wage subsidies	0.5		2.7
Lead(t+5) disabled wage subsidies	0.5		3.8
Lead(t+3) Full DI		12.1	32.1
Lead(t+5) Full DI		16.9	41.9

Note: Full DI refers to those awarded DI. Disabled wage subsidies refer to those awarded disabled wage subsidies (rejected for full DI). Rejected refers to those who were rejected for full DI, and were not awarded disabled wage subsidies.

From table 2 we see the following important points: a) those awarded full DI and those rejected were both less employed after the application than in t-1, while those awarded a disabled wage subsidies were more employed than in t-1, b) those awarded DI hardly worked, even though they are allowed to work[57], c) those awarded a disabled wage subsidies were employed a large share of the year even 5 years out (66.7%)[58], d) those rejected for full DI (who did not receive a disabled wage subsidies) only worked 23.8% of the year 5 years after they were rejected, e) very few of those awarded full DI changed to disabled wage subsidies 5 years out (0.5%), f) 16.9% of those awarded disabled wage subsidies received

57 For each hour they work the loose one hour of benefit.

58 Individuals awarded disabled wage subsidies will receive unemployment benefit and sickness benefit, when they lose their jobs or when they are sick from work.

full DI 5 years out, g) 41.9% of those rejected for full DI (who did not receive disabled wage subsidies) received full DI 5 years after, while only 3.8% received disabled wage subsidies.

The most important finding from table 2 is the big difference between those rejected for full DI and not receiving a disabled wage subsidies and those receiving a disabled wage subsidies. Those receiving a disabled wage subsidies worked 3 times as much, 5 years after applying for DI, as those who were "just" rejected.[59] It might not be surprising that a lot of those receiving disabled wage subsidies worked since they per definition were secured some kind of work; it is more surprising that such a large percentage of those who were "rejected only" did not work.

4.3 Results from audit reports

The audit reform in 1995 aimed at harmonizing the award decisions across municipalities through an audit of a random sample of awards in a rolling two counties every year from 1995 to 2001. State lawyers re-assessed around 100 decisions in each county each year and subsequently reported how many of the rulings they agreed on, and they did not agree on. These reports allow us to differentiate between the number of cases where municipalities rejected DI and the state lawyers awarded DI (Type 1 errors) and those cases where municipalities awarded DI and the state lawyers rejected DI(Type 2 errors). In table 3 we present the main results from the audit reports.

Table 3: Type 1 and type 2 errors from state audits of DI awards, 1995-2001

Year audited	County	Number of cases	Type 1	Type 2	Type 1 %	Type 2 %
1995	Storestrøms amt Ringkøbing amt	Missing	Missing	Missing	Missing	Missing
1996	Københavns amt	99	0	7	0.0	7.1
	Sønderjylland amt	97	1	18	1.0	20.9
1997	Ribe amt	99	0	19	0.0	21.6
	Nordjyllands amt	106	0	15	0.0	14.9
1998	Århus amt Københavns and Frederiksberg kommune	Missing	Missing	Missing	Missing	Missing
1999	Storstrøms amt	72	0	1	0.0	1.4
	Sønderjylland amt	69	0	2	0.0	3.0
2000	Fredriksborg amt	81	0	4	0.0	5.4
	Bornholm amt	25	0	0	0.0	0.0
2001	Viborg amt Vejle amt Roskilde amt	Missing	Missing	Missing	Missing	Missing

Note: A type 1 error is a case where a county do not award DI but the National Social Appeals board does, and a type 2 error is the opposite case. It has not been possible to acquire the reports from the 1995, 1998 and 2001–the Social Appeals board does no longer hold these reports.

[59] This finding is in good agreement with the findings in Wachter et al. (2008).

Three out of the seven reports are no longer available from the National Social Appeals board. Therefore, we have to exclude those counties for which we do not have this information when we analyze the effect of the reform. From table 3 we see that there were almost no type 1 errors, where the municipality rejected and the appeal board awarded DI, but there were rather many type 2 errors where the appeal board rejected and county awarded DI. This indicates that the reform would increase rejection rates, because counties would have to reject more applicants in order to meet best practice standards from the National Social Appeals Board. The column "Type 2 %" shows that there was a lot of variation in the deviation from best practices between counties and over time. If we, for example, compare type 2 errors between the two counties audited in 1995, then "Storestrøms amt" had 37% categorized as type 2 errors while "Ringkøbing amt" only had 1%. Generally, the tendency is that the type 2 errors go down over time, which could indicate that the counties are starting to adapt best practices.

5. Estimation and results

Difficulties in observing an applicant's true health (for the DI awarding authorities) encourages some individuals to apply for DI even though they are capable of working. This has motivated governments in many countries to reform their DI programs to better reveal the health of the applicants. Typically two strategies have been pursued: stricter screening or reduction of benefits. We exploit different stringency reforms of the Danish DI system in the 1990' which all tried to mitigate moral hazard.

The effects on municipal awards of the introduction of state audits in 1995 and the reduction in rebates in the 1997 reform will be analyzed in reduced form by difference-in-differences. A first set of difference-in-difference analyses use year-region-specific audit decisions regressed on changes in award probabilities from t-1 to t. A second set of difference-in differences equations analyzes how reduced rebates to the age-group 60-66 affected changes in award probabilities over time for different age groups. In a final set of instrumental variables regressions, we use municipality-year specific variations in award propensity to determine individual award probabilities in a first stage regression to identify a LATE from a second stage individual labor supply equation. Simulations allow us to illustrate how each stringency policy tool affects award probabilities and thereby differentially affects labor supply decisions.

5.1 Difference-in-Difference estimation of the 1995 audit reform

The 1995 audit reform aimed at harmonizing best practice in DI awards, by letting the National Social Appeal board audit about 200 cases from two counties each year from 1995 to 2001. The audit reform can be seen as a natural experiment, where two out of 15 counties were randomly picked for auditing (the treatment group) and the other 13 counties constituted a natural control group. We exploit the random assignment and the natural control group design to identify the average reduction in rejection rates due to the auditing.[60]

The 15 counties were divided into 275 municipalities that each awarded DI pensions. We therefore estimate our reduced form model at the municipality year level. We expect the audits primarily to have an effect the year that the audit reports were published, which normally happened at the beginning of the year following the audit. We know what kind of disagreement (type 1 or type 2 error) the audit

---

[60] Since a larger share of the control group is audited over time, we expect our estimate to be underestimating the true effect of the auditing, because the average rejection rate would fall over time in the control group too. We also assume that rejection probabilities did not make individuals move between municipalities in order to increase the probability of being awarded DI.

reports found, and we expected different effects on the awarding process depending on whether it was a type 1 (municipality too stringent in their awarding practice) or if it was a type 2 error (municipalities were too lenient their awarding practice). If a type 1 error was observed, then we expected the municipalities to lower their rejection rate subsequently, while we expected an increase in the rejection rate from a type 2 error. As we saw in the descriptive section, then almost all of the disagreements between the National Social Appeal board and the awarding authorities in the municipalities were type 2 errors (municipalities awarded to many DI pensions), so we expected the auditing reform to have increased rejection rates, but we still allowed for separate effects of the two types of errors. We restricted our estimation period to 1993-2001 for two reasons. First we regressed the type 1 and type 2 errors on changes in rejection rates over time in each municipality-year, which required us to drop the first year 1992. Secondly since all 15 counties were to be audited over time, then the random element in the selection of treatment groups diminishes over time, therefore we restricted our estimation until 2001 and drop 2002, since five counties still needed to be checked by the beginning of year 2000 (we measure the effect the next year). We estimate the following regression equation controlling for different X's to test robustness of the estimate:

$$(1) \qquad \Delta(reject)_{jt} = \alpha + \beta_1 * Type1_{jt} + \beta_2 * Type2_{jt} + \beta_{3-12} * Year_t + \beta_{13-288} * Muni_j + \beta_{289-Z} X_{jt} + e_{jt}$$

$\Delta reject_{jt}$ is the change in average rejection rate in each municipality from t-1 to t, Type1 and Type 2 are the average share of audited cases that the municipality and the appeal board disagreed on. Year is a set of dummies that captures business cycle effects and other year specific time effects. Muni is a dummy that captures time invariant municipality effects. $X_{286\text{-}z}$ contains different sets of controls depending on which model we estimate.[61]

Table 4 show the result of the estimation of equation (1) with different sets of controls, to check the robustness of the effect. The coefficient on the type 1 error is negative as expected but highly insignificant–that might be because there were almost no Type 1 cases, cf. table3. The coefficient on the type 2 error variable is positive as expected and almost significant at the 1% level in all models.

[61] We have chosen the controls based on the investigation in Weatherall (2002), which investigates what factors that increase the probability of receiving DI.

Table 4: Difference-in-Difference estimation 1995 audit reform

	Model 1	Model 2	Model 3	Model 4
	margins_b / P > |z|	margins_b / P > |z|	margins_b / P > |z|	margins_b / P > |z|
Type 1 error	-0.0327	-0.0316	-0.0307	-0.0309
	(0.260)	(0.275)	(0.290)	(0.287)
Type 2 error	0.2160**	0.2170**	0.2140**	0.2135**
	(0.010)	(0.010)	(0.011)	(0.011)
9 Year fixed effect	x	x	x	x
275 Municipality fixed effects	x	x	x	x
Average age municipality-year		x	x	x
Share single municipality-year		x	x	x
Income municipality-year		x	x	x
Share men municipality-year			x	x
Share with basic school municipality-year			x	x
Share general upper secondary school municipality-year			x	x
Share receiving cash pay municipality-year			x	x
Share lowest level employees municipality-year			x	x
Share with no diagnoses municipality-year				x
Share of time sick municipality-year				x
Average days in hospital municipality-year				x
Observations	2,282	2,282	2,282	2,282
R-square	0.043	0.046	0.049	0.050

Note: Type 1 error refers to cases where the municipality rejected and the Social Appeals board awarded, and type 2 represent the opposite outcome. The p-values refer to the marginal effect of a Type 1 and a Type 2 error. We weight each municipality year according to number of applicants in each cell.

The positive marginal effect on a Type 2 error indicates that the average municipality-year rejection rate increased by 0.21% on average for every 1% increase in disagreement between municipalities and the audit board. The average type 2 error was 9.7%, which means that the average rejection rate would increase by 2 percentage points if the average type 2 error was reduced to 0. Table 3 show that "Ribe amt" was the county showing the largest disagreement with the audit board (Type 2 = 21.6%). If "Ribe amt" could bring the disagreement down to 0%, then we would expect the rejection rate to increase by 4.5 percentage points in "Ribe amt".

From table 4 we conclude that the audit reform managed to increase rejection rates for those municipalities that awarded too many DI pensions prior to the audit reform.

5.2 Difference-in-Difference estimation of the 1997 rebate reform

The state rebate to the municipalities for DI expenditures on benefits for those aged 60-66 was 100% until 1997. The reform reduced the rebate to 50% to encourage municipalities to award fewer DI benefits for this age group. The rebate rate was already 50% for those aged 16-59 before the reform. The main difference between this reform and the audit reform that we considered in section 5.1 is that the 1997 rebate reform tried to motivate fewer awards through an financial penalty (or removal of subsidy) to the municipalities, whereas the 1995 audit reform encouraged best practices through auditing. Rejection rates for those aged 60-66 were lower before the reform; therefore, we compare the difference in rejection rates between those aged < 60 and those aged 60-66 before and after the reform. We test whether this difference was significantly reduced after the reform. We estimate the difference-in-difference at the municipality year age group level. The equation we estimate is defined in the following way:

$$(2)\ reject_{jta} = \alpha + \beta_1 * d60before_{jta} + \beta_2 * d60after_{jta} + \beta_{3-12} * Year_t + \beta_{13-288} * Muni_j + \beta_{289-Z} * X_{jta} + \varepsilon_{jta}$$

$Reject_{jta}$ is the average rejection rate in each municipality year by age group, d60before measures the rejection rate difference before the reform between those aged < 60 and those aged 60-66, d60after measures the rejection rate difference after the reform between those aged < 60 and those aged 60-66. Year is a fixed effect that captures business cycle effects and other year-specific time effects. Muni is a dummy that captures time invariant municipality specific effects. $X_{288\text{-}z}$ contains a different set of controls depending on which model we estimate. When we estimate the model, we exclude 1996 from the estimation, to avoid Ashenfelter's dip.[62]

Table 5 shows the result of the estimation of equation of (2), with different sets of controls. From table 5 we see that both dummies are negative and significant in all models, indicating that those aged 60 on average had a lower rejection rate than those aged < 60. A second point to notice is that the after reform dummy (DUM_1997_2002_60_66) is less negative than the before reform dummy, indicating that the difference between the average rejection rate of those aged 60+ and those aged < 60 decreased after the reform. The difference between the two dummies is fairly stable across models–between 0.046 and 0.072–and the F-test shows that the difference between the two dummies is significant below the 1% level.

From the results we conclude that the average rejection rate for those aged 60+ was reduced by 5 percentage points (evaluated by the model with a full set of controls) compared to those aged < 60 after the rebate reform was implemented in 1997.

---

[62] This is the phenomenon whereby an outcome or selection factor falls or dips before program entry. Not accounting for this would exaggerate program effects (Ashenfelter, 1978).

Table 5: Difference-in-Difference estimation 1997 rebate reform

	Model 5	Model 6	Model 7	Model 8
	margins_b / std.	margins_b / std.	margins_b / std.	margins_b / std.
DUM_1992_1995_60_66	-0.0619***	-0.0867***	-0.0791***	-0.0497***
	(0.0043)	(0.0058)	(0.0057)	(0.0052)
DUM_1997_2002_60_66	-0.0111***	-0.0147***	-0.0147***	-0.0034**
	(0.0006)	(0.0010)	(0.001)	(0.0011)
Reduction in difference between before and after	0.051	0.072	0.064	0.046
(P > F) : DUM_1992_1995_60_66 = DUM_1997_2002_60_66	0.001	0.000	0.000	0.003
9 Year fixed effect	x	x	x	x
275 Municipality fixed effects	x	x	x	x
Share single municipality-year-age		x	x	x
Income municipality-year-age		x	x	x
Share men municipality-year-age			x	x
Share with basic school municipality-year-age			x	x
Share general upper secondary school municipality-year-age			x	x
Share receiving cash pay municipality-year-age			x	x
Share lowest level employees municipality-year-age			x	x
Share with no diagnoses municipality-year-age				x
Share of time sick municipality-year-age				x
Average days in hospital municipality-year-age				x
Observations	2,861	2,861	2,861	2,850
R-square	0.357	0.374	0.389	0.457

Note: The F-test test whether the difference in the average rejection rate level between those aged 60+ and those aged < 60 changed from before to after the reform in 1997. The estimation is restricted to cells with at least 3 cases. We weight each municipality-year-age group according to number of applicants in each cell.

5.3 Labor supply of rejected DI applicants

Many have analyzed how DI benefits give rise to moral hazard. They have often done so by looking at how a high replacement rate distorts labor supply. Since replacement rates for DI are decreasing functions of past earnings, it is difficult to determine whether it is generous replacement rates or low earnings that induce individuals to leave the labor force. Bound (1989) was the first one to criticize the replacement rate approach for having an endogeneity problem, and instead used rejection rates to determine labor supply for those rejected. Bound (1989) argued that those rejected for DI formed an upper bound (because those rejected are in slightly better health than those awarded DI) on the expected labor supply to be gained from rejecting one more individual. Gruber and Kubik (1997) expanded the insight of Bound (1989) by using a natural experiment that provided variation in rejection rates across states, and they conducted a reduced form difference-in-differences analysis. Gruber and Kubik (1997) did not have information on individual rejections, so they estimated a reduced form labor supply equation with state-year rejection rates.

We have information on individual rejections and exploit this in a first stage regression by instrumenting individual rejections by average municipality year rejection rates. Since we have information about individual rejection rates (which Gruber and Kubik did not), we can estimate local average treatment effects by instrumental variables, by exploiting within municipality variation in award probabilities. Based on the difference-in-difference analysis of the 1995 audit reform and the 1997 rebate reform and the descriptive statistics of the development in rejection rates after the 1998 filtering reform, we expect most of the changes in municipality year rejection rates to be due to these reforms, since there were no changes to DI benefit generosity in Denmark during our estimation period (1992-2002).

The 1998 reform coincided with an expansion of a wage subsidy program for the partially disabled—disabled wage subsidies. Disabled wage subsidies is similar to full DI in the sense that applicants have to be permanently disabled to be considered for the program and both programs are permanent in that they are not subject to re-assessment of eligability. But there was one important distinction—people receiving disabled wage subsidies have to work.

Since many factors affect an individual's rejection status, there is a big chance that rejection is correlated with omitted variables and therefore endogenous. Omitted variables could be preferences for leisure or lack of motivation to work. Both variables would cause the rejection rate effect on the labor supply to be underestimated, because individuals rejected for DI might substitute to other welfare programs instead of applying for work. We later test if the rejection rate is endogenous.
To deal with this problem, we instrument the individual rejection by the average rejection rate in each municipality-year age group, and use the exogenous variation that the reforms created in awarding across municipalities to identify the effect on the labor supply. We also add municipality and year fixed effects, municipality time varying effects and individual controls. The first stage equation is defined as:

$$(3)\quad reject_{ijt} = \alpha + \beta_1 * reject\_mya_{jta} + \beta_{2-11} * Year + \beta_{12-287} * Muni\_1_j + \beta_{288-290} * Muni\_2_{jt} + \beta_{291-299} * X_{ijt} + \varepsilon_{ijt}$$

$Reject_{ijt}$ is the individual rejection rate in year t, $reject\_mya_{jta}$ is the average municipality-year age group rejection rate. $Year_t$ and $Muni1_j$ are the fixed time and municipality effects, and $Muni2_{jt}$ is the time varying municipality effects (average employment in municipality, average time in rehabilitation in municipality and average time receiving cash benefits in municipality). X_{ijt} is a vector of individual controls (age, gender, ethnicity, marital status, days in hospital, days sick from work, socioeconomic status, education level, and medical diagnosis).

In the second stage we estimate labor supply as a function of the instrumented individual rejection rate in order to capture the effect of the three DI reforms on the individual labor supply. We measure labor supply one year after DI application. We measure labor supply by the percentage of the year an applicant was working according to the DREAM[63] database. We estimate the following labor supply equation:

$$(4)\quad LS(t+1)_{ijt} = \alpha + \beta_1 * reject\_hat_{ijt} + \beta_{2-11} * Year + \beta_{12-287} * Muni\_1_j + \beta_{288-290} * Muni\_2_{jt} + \beta_{291-299} * X_{ijt} + \varepsilon_{ijt}$$

Reject_hat$_{ijt}$ is the instrumented individual rejection rate from the first stage, and the rest of the variables are the same as those described in equation (3).

Table 6 shows the results from estimating the first stage equation (equation 3) and the second stage labor supply equation (equation (4)) for the age group 30-66 over the period 1993-2002. The models are estimated with the full set of controls described in equation (3) and (4), and only with municipality and year fixed effects, to check the robustness of the instrument and the instrumented individual rejection rate.

From the first stage we see that the instrument (rejection rate in municipality year age group) is highly significant in both the simple model (model9) and the model with a full set of controls (model 10), but the marginal effect decreases from 0.34 to 0.19 after the controls are added. This shows that it is important to control for demographic factors and fixed and time effects in each municipality-year-age group, in order not to overestimate the effect of the instrument. Even though we control for all these factors, the rejection rate in municipality-year-age group remains highly significant, which suggests that there also exists sufficient exogeneous variation in the average rejection rate across municipalities. The average increase in individual rejection rates from increasing municipality rejection rates by 1% is 0.19% in the model with a full set of controls (model 10).

63 DREAM contains information on labor market status each week. See the data section for a description of DREAM.

Table 6: First and second stage regressions, age 30-66, 1993-2002

	First stage		Second stage	
Dependent variable	Individual rejection rate		Employment t+1	
	Model 9	Model 10	Model 11	Model 12
	margins_b / std.	margins_b / std.	margins_b / std.	margins_b / std.
Instrument: rejection in municipality-year-age group	0.3395***	0.1904***		
	(0.0084)	(0.0092)		
Instrumented individual rejection rate			0.5862***	0.6419***
			-0.0235	-0.0455
Year fixed effects		x		x
Municipality fixed effects		x		x
Dummy for diagnosis		x		x
Age groups		x		x
Gender		x		x
Single		x		x
Education (5 levels)		x		x
Days in hospital (t-1)		x		x
Days sick (t-1)		x		x
Days receiving rehabilitation (t-1)		x		x
Days receiving cash pay (t-1)		x		x
Average days in hospital municipality year		x		x
Average days sick municipality year		x		x
Average income in municipality year		x		x
Observations	61,858	60,663	59,179	58,085
Adjusted R2	0.078	0.128	0.477	0.52

Note: When we calculate rejection rates in municipality year we exclude the individual itself from the calculation.

The second stage shows that the instrumented individual rejection rate is highly significant in both the simple model (model 11) and in the model with a full set of controls (model 12). The marginal effect is estimated to 0.64 meaning that a 100% increase in individual rejection rates would increase employment by 64% on average. The results suggest that variations in rejection rates across municipalities are important for explaining variations in labor supply across municipalities.[64]

In table 2 we showed that those who were awarded disabled wage subsidies (rejected for full DI) on average were employed 77.8%[65] of the year in t+1 while those rejected for full DI and who did not get

[64] If we estimate the reduced form model directly by OLS using the observed individual rejection rates instead of the instrumented individual rejection rate, we estimate β_{OLS} =0.564. In the IV-regression in table 6 β_{IV} =0.423. If we use the Durbin-Wu-Hausman test to test, if reject is endogenous we find an $F\text{-value}_{(1,57,780)} = 22.35$, with p=0.0000. We therefore conclude that reject is indeed endogenous.

[65] Those in disabled wage subsidies can also be unemployed, receive sickness benefit and lose their disabled wage subsidies if they job ends.

a disabled wage subsidies were only employed 34.1% of the year in t+1. This difference in employment between the two groups increased by t+5 to 66.7% vs. 23.8%.

From these results we conclude that increased rejection rates increased employment, and those who were rejected for full DI and awarded a disabled wage subsidies supplied more labor in the short and long run than those rejected and not awarded a disabled wage subsidies.

5.4 Did the reform manage to target the right individuals?

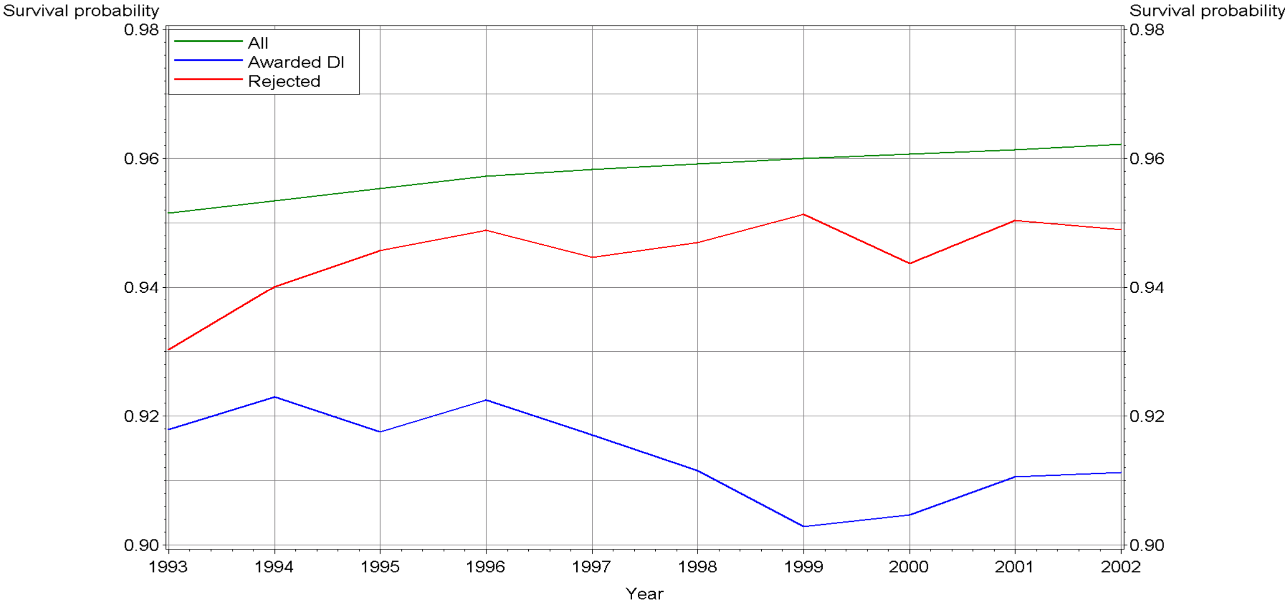
Our estimations so far have shown that the three stringency reforms managed to increase rejection rates and labor supply. To evaluate whether the reforms succeeded in rejecting the healthiest applicants and not the sickest applicants, we calculate the 5-year survival probability conditioning on demographics and health. We compare the development in survival rates between the population (baseline), those awarded DI and those rejected for DI (disabled wage subsidies awards + other rejected).[66] Generally we expect an overall increase in survival probabilities because of the increasing trend in life expectancies, but we expect the increase to be lower (or negative) for those awarded DI, because that would indicate that increased stringency resulted in fewer awards to the healthiest.

Figure 5 shows the development in 5-year survival probabilities over time (1993-2002) for the population, those awarded DI and those rejected for DI (disabled wage subsidies awards + other rejected), for men aged 50-59.

Figure 5 shows that the survival probability for those awarded DI fell after 1996, whereas the survival probability for the population increased over the entire period. The survival probability for those rejected was almost the same for the period after 1996. This indicates that increased stringency meant rejecting those applicants in the best health. Figure 1a-3a in Appendix A shows the same figure for women aged 50-59 and men and women aged 40-45. The general picture is the same. From this we conclude that the increased stringency primarily rejected the healthiest applicants.

[66] We assume that being awarded full DI, disabled wage subsidies or rejected did not affect the survival probability.

Figure 5: Survival probability 5 years out for the population, those awarded DI and rejected, age group 50-59, men



NOTE: Rejected include those awarded disabled wage subsidies.

6. Conclusion

There has been an increase in DI recipients over the last 3-4 decades, despite the fact that general health and life expectancy have been improving. This has made researchers argue that generous benefits, combined with difficulties in defining and observing disability, create moral hazard problems that distort labor supply.

The two broad policy approaches to tackle the moral hazard problem has been improved targeting of awards through increased stringency and reduction of benefits. We use stringency reforms that directly targeted the screening process of the DI awarding process to estimate how increased rejection rates affect labor supply.

We first measure the effect of two of the three[67] stringency reforms (1995 audit reform and the 1997 rebate reform) in a difference-in-difference setup. From the DID analysis of the 1995 audit reform we find that a 10% increase in disagreement (municipalities awarded to many DI's) about awards between the municipality and the National Social Appeals board increased rejection rates by 2.1% the following year when the report was published. From the DID analysis of the 1997 rebate reform we find that rejection rates decreased on average 4.6 percentage points in the age group 60+ (treatment group–rebates reduced from 100% to 50%) compared to the age group < 60 (control group–rebates 50%).

From the two stage labor supply equation estimation we find that municipality year rejection rates are a strong instrument for individual rejection rates even when controlling for demographics and health information–the average increase in individual rejection rates from increasing municipality year rejection rates by 10 is 1.9%. The second stage shows that the instrumented individual rejection rate is highly significant in explaining future employment–a 100% increase in individual rejection rates would increase employment by 64% on average in t+1. Those who were awarded a disabled wage subsidies (rejected for full DI) were on average employed 77.8% of the year in t+1 while those rejected for full DI and who did not get a disabled wage subsidy were only employed 34.1% of the year in t+1. From these results we conclude that an increase in rejection rates increased employment, and those who were rejected for full DI and awarded a disabled wage subsidy supplied more labor in t+1 than those who were rejected and not awarded a disabled wage subsidies.

---

[67] The last reform (1998 filtering reform) does not have a natural control group and can therefore not be analyzed in a DID setup.

Finally we investigate whether the increased stringency, which resulted in an increase in rejection rates, managed to target the healthiest applicants. We measure this by comparing survival rates 5 years out. We find evidence that those awarded DI experienced a lower increase in survival probability, indicating that the increased stringency managed to reject the healthier applicants.

Appendix

Appendix A

Figure 1a: Survival probability 5 years out for the population, those awarded DI and rejected, age group 50-59, women

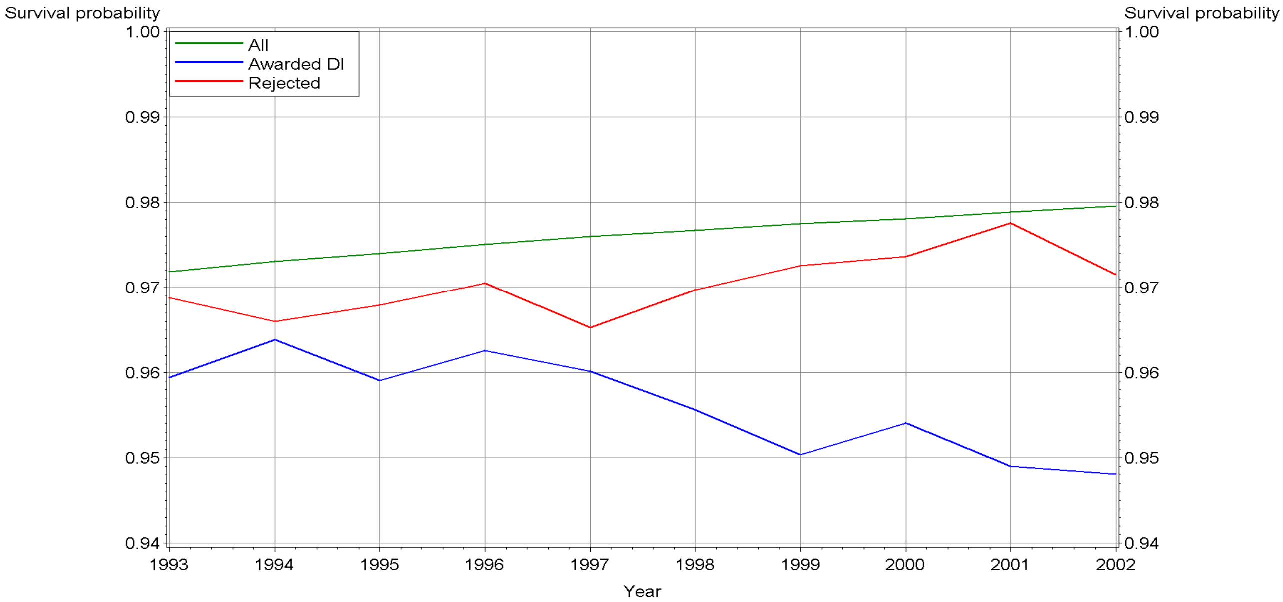


Figure 2a: Survival probability 5 years out for the population, those awarded DI and rejected, age group 40-45, men

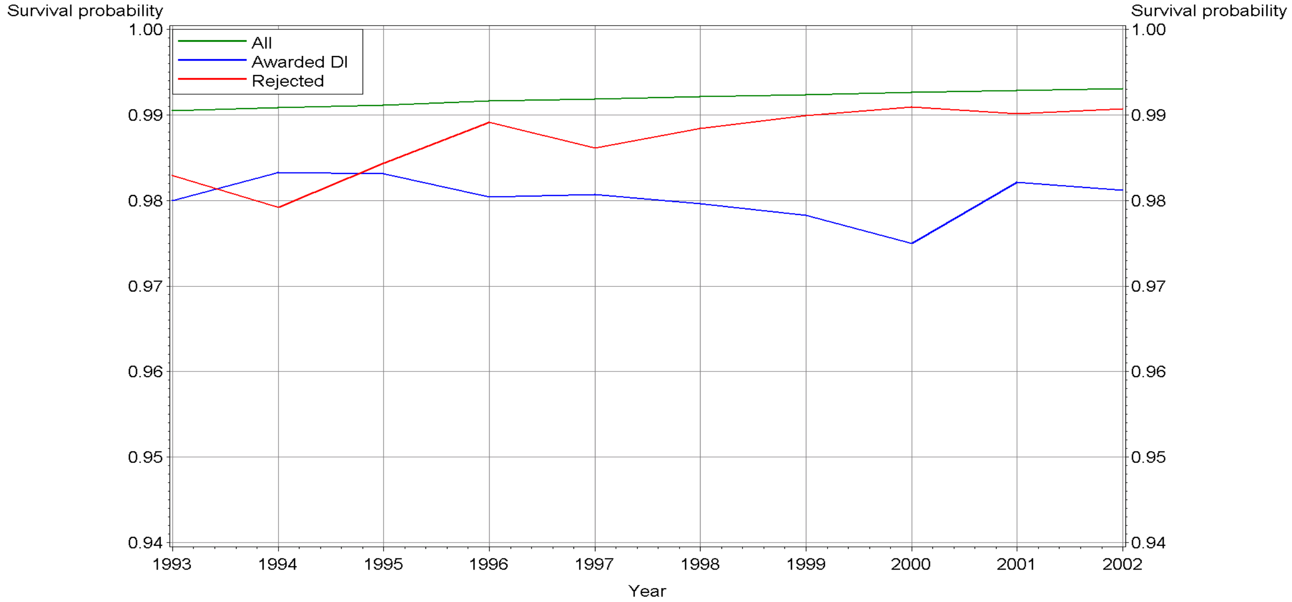
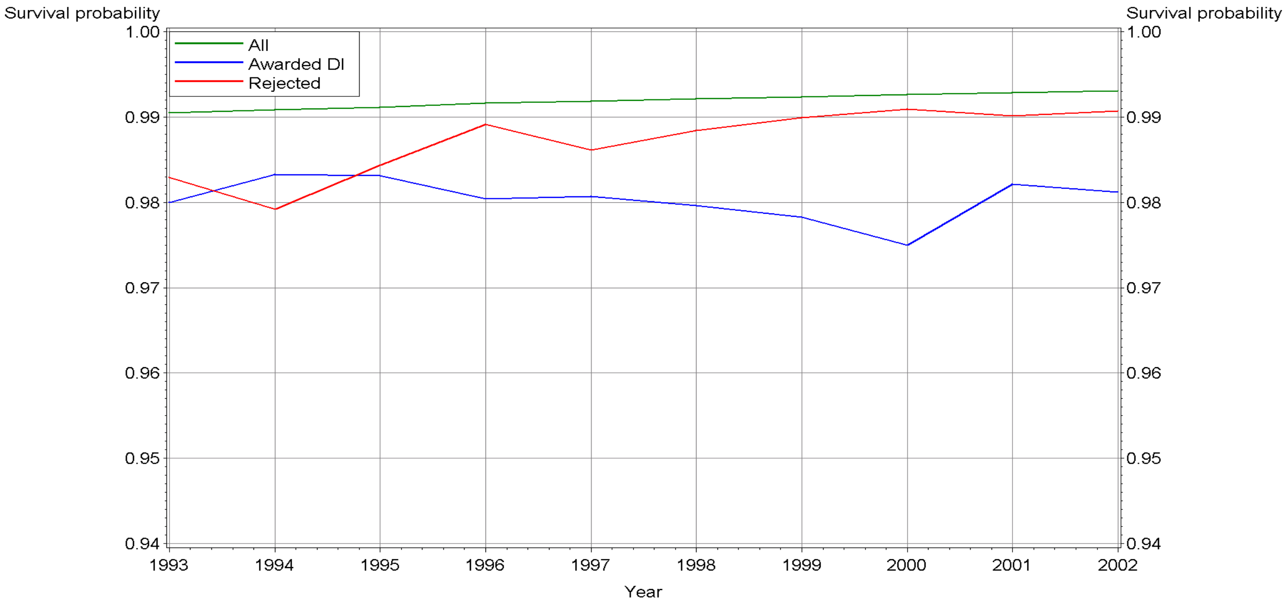


Figure 3a: Survival probability 5 years out for the population, those awarded DI and rejected, age group 45-49, women



References

Ashenfelter O (1978): "Estimating the Effect of Training Programs on Earnings". *Review of Economics and Statistics*, Vol. 60, Issue 1, pp. 47-57.

Autor DH and Duggan MG (2003): "The Rise in the Disability Rolls and the Decline in Unemployment". *The Quarterly Journal of Economics*, Vol. 118, No. 1, pp. 157-205.

Autor DH and Duggan MG (2006): "The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding". *Journal of Economic Perspectives, American Economic Association*, Vol. 20, No. 3, pp. 71-96.

Autor D, Maestas N, Mulen K, and Strand A (2011): *Does Delay Cause Decay? The Effect of Administrative Decision Time on Labor Force Participation and Earnings of Disability Applicants.* Working Paper 2011-258. Michigan Retirement Research Center.

Bengtsson S (2002): *Bestemmer forvaltningen om du får førtidspension? Kommunens forvaltningspraksis og tilkendelse af førtidspension*. Socialforskningsinstituttet, 02:15.

Benitez-Silva H, Buchinsky M, Chan HM, Rust J, and Sheidvasser S (1999): "An Empirical Analysis of the Social Security Disability Application, Appeal and Award Process". *Labour Economics*, Vol. 6, No. 2, pp. 147-178.

Borghans L, Gielen AC, and Luttmer E (2012): *Social Support Substitution and Labor Supply Shifting.* Mimeo.

Bound J (1991): "Health and Earnings or Rejected Disability Insurance Applicants: Reply". *American Economic Review*, Vol. 81, No. 5, pp. 1427-1434.

Bound J (1989): "Health and Earnings or Rejected Disability Insurance Applicants". *American Economic Review*, Vol. 79, No. 3, pp. 482-503.

Bound J and Burkhauser RV (1999): "Economic Analysis on Transfer Programs Targeted on People with Disabilities". In Ashenfelter O and Card D (Eds.): *Handbook of Labor Economics*, Vol. 3c, Elsevier, pp. 3417-3528.

Chen S and Klaauw W van der (2008): "The Work Disincentive Effects of the Disability Insurance Program in the 1990's". *Journal of Econometrics*, Vol. 142, No. 2, pp. 757-784.

Danish Economic Council (2003): "Aktiv socialpolitik og førtidspension". In: *Dansk Økonomi. Forår 2005. Konjunkturvurdering. Indkomstoverførsler og velfærdsstaten*. Det økonomiske råd. Formandskabet.

De Jong P, Lindeboom M, and Klaauw, B. van der (2011): "Screening Disability Insurance Applications". *Journal of the European Economic Association*, Vol. 9, No. 1, pp. 106-129.

French E and Song J (2011): *The Effect of Disability Insurance Receipt on Labor Supply*. Mimeo.

Gruber J (2000): "Disability Insurance Benefits and Labor Supply. *Journal of Political Economy*, Vol. 108, No. 6, pp. 1162-83.

Gruber J and Kubik JD (1997): "Disability Insurance Rejection Rates, and the Labor Supply of Older Workers". *Journal of Public Economics*, Vol. 64, No. 1, pp. 1-23

Karlström A, Palme M, and Svensson I (2008): "The Employment Effect of Stricter Rules for Eligibility for DI: Evidence from a Natural Experiment in Sweden". *Journal of Public Economics*, Vol. 92, pp. 2071-2082.

Maestas N, Mulen K, and Strand A (2011): *Do Disability Benefits Discourage Work?* Mimeo.

Parsons DO (1991): "Health and Earnings or Rejected Disability Insurance Applicants: Comment". *American Economic Review*, Vol. 81, No., pp. 1419-1426.

Parson DO (1980): "The Decline in Male Labor Force Participation". *Journal of Political Economy*, Vol. 88, No. 1, pp. 117-134.

Wachter T von, Song J, and Manchester J (2008): *The Employment Effects of Social Security Disability Insurance in the Past 25 Years: A Study of Rejected Applicants Using Administrative Data.* Discussion Papers 0809-05, Columbia University, Department of Economics.

Wachter T von, Song J, and Manchester J (2011): "Trends in Employment and Earnings of Allowed and Rejected Applicants to the Social Security Disability Insurance Program". *American Economic Review*, Vol. 101, Issue 7, pp. 3308-3329.

Weatherall JH (2002): *Vejen til førtidspension. En analyse af overgangen til førtidspension i befolkningen*. Socialforskningsinstituttet, 02:9.