Sickness absence and voluntary employer paid health insurance

Kjeld Møller Pedersen

Institute of public health, unit of health economics research

University of Southern Denmark

kmp@sam.sdu.dk

Table of Contents

Abstract

Sickness absence is a problem with considerable economic dimensions. About 4% of the total annual working days are lost due to absence. Therefore, ways to reduce absence are eagerly sought. In a Danish context employer paid insurance is but one example. The tax exempt status of this type of voluntary duplicate health insurance has been argued by reference to the potential for reducing long term sickness absence. However, nationally and internationally there is no evidence about this. The present paper analyzes this theoretically and empirically. A simple model for 'demand' for sickness absence in the Grossman-tradition is used. Empirically, two recent survey data sets are used. The determinants of absence are analyzed using quantile regression in order to look at the extreme parts of the conditional distribution, e.g. 90% og 95% for long term absence. No significant results are found on the absence reducing property of health insurance. A two component ('short' and 'long' term absence) finite mixture model is also applied with the same result. The problems with a causal interpretation of regression analyses may (partly) be circumvented by using (correctly specified) propensity scores and matching estimators. Regression analysis and propensity score, however, share the same challenge: Both are based on selection based on observables. Using the matching estimator approach there are no signs of a treatment effect of health insurance using the presenteeism data set, while there is evidence using the health insurance data set. However, the specification of the propensity score for the latter is not as exhaustive as for presenteeism data set, and in some cases there are statistically significant differences for some control variables after matching.

JEL: J22, I12

Keywords: absenteeism, voluntary health insurance

Introduction[I]

When the current Danish legislation on employer paid health insurance for employees was enacted mid 2002, one of the main arguments for tax exemption of this particular employee benefit[II] was that it was expected that it would reduce long term sickness absence[III]. One of the government's supporting arguments for the legislation went as follows: "… it is an advantage for the employer, who will see reduced sickness absence and faster return to work, hence avoiding costs of both economic and organizational nature associated with long term employee absence"[1].

There is no tradition for providing empirical evidence for such (political) statements. It is either 'common belief' or a politically expedient type of argument. However, post festum of enactment it is always of interest to investigate empirically whether the claims hold up to scrutiny. The present work is such an analysis based on two available data sources[IV].

It is obvious that evidence on this type of effect of voluntary duplicate (VD) health insurance[V] is relevant not only for policy purposes but also in general in relation to the empirical literature on the effects of VD health insurance. A quick perusal of the existing theoretical and empirical literature reveals very few studies on the effect on sickness absence – in reality only a Danish study and a working paper draft[4, 5] along with a working paper from 1985 co-authored by the present author[70]. It is important to stress that the focus here is on *VD health insurance* and sickness absence, and *not* the effects of various kinds of *sickness absence insurance* on the absence rate. Most of the

[I] Funding from Helsefonden for data collection is gratefully acknowledged. Jacob Nielsen Arendt and Astrid Kiil, University of Southern Denmark, have provided good, concise, and very useful comments. They also were partners in the health insurance survey, one of the data sources used in the present paper. Useful comments have been received from Lars Skipper, discussant at a seminar organized by the Danish Insurance Association and from Nabanita Datta Gupta, discussant, at the annual 2011 meeting of the Forum for Danish Health Economists. Comments from participants at an in-house departmental seminar are also acknowledged.

[II] In Denmark most employer provided fringe benefits, e.g. 'free telephone' or 'free company car', are subject to income taxation based on an imputed value of the fringe benefit in question. Thus, somewhat unusual the employer paid health insurance was exempted provided that all employees of a company were offered the insurance.

[III] See table A, appendix I, for stated reasons for holding health insurance among insured employees. Reduction of sickness absence was at the top of the list.

[IV] Recently the National Audit Office/GAO (Rigsrevisionen) has asked the Ministry of Health to document the effect of employers paid health insurance on the waiting time for treatment. This was another of the arguments for tax exemption.

[V] In the health insurance literature is common to distinguish between complementary, supplementary or duplicate health insurance in relation to the tax-financed system[2, 3]: *1*. Complementary voluntary private health insurance covers co-payments for treatments that are only partly covered by the tax-financed health care system. *2*. Supplementary voluntary private health insurance covers treatments that are excluded from the tax-financed health care system. *3*. Duplicate voluntary private health insurance covers diagnostics and elective surgery at private hospitals and for instance physiotherapy or office visits to medical specialists. – services that are also provided by the tax-financed public health care system.

economic literature on sickness absence is about the latter issue, and hence the effect on sickness absence of degree of economic compensation.

The aims of the present study are, first of all, to estimate the possible effect of health insurance on sickness absence, both short term (one to several days of sporadic absence during the year) and long term (spells of >15 consecutive days of absence)[VI], secondly, as a necessary prerequisite for the first question, to survey briefly the relevant (economic) literature and development of a model.

The remainder of this paper is organized as follows. In the background section the Danish situation as regards sickness absence and health insurance is briefly sketched. This is followed by a section on theoretical background in which a simple model for 'demand' for absence is presented. This is followed by a description of the data and a section with a few descriptive results. The statistical analysis consist of a section where quantile regression is used to study the whole (conditional) distribution of sickness absence to distinguish effects on short- and long-term absence – apparently the first use of this approach in sickness absence research - and a section with propensity score and matching estimators to estimate mean effects and in an attempt to get closer to a causal interpretation of health insurance's possible influence on sickness absence. The closing section provides a discussion of results and perspectives.

Background

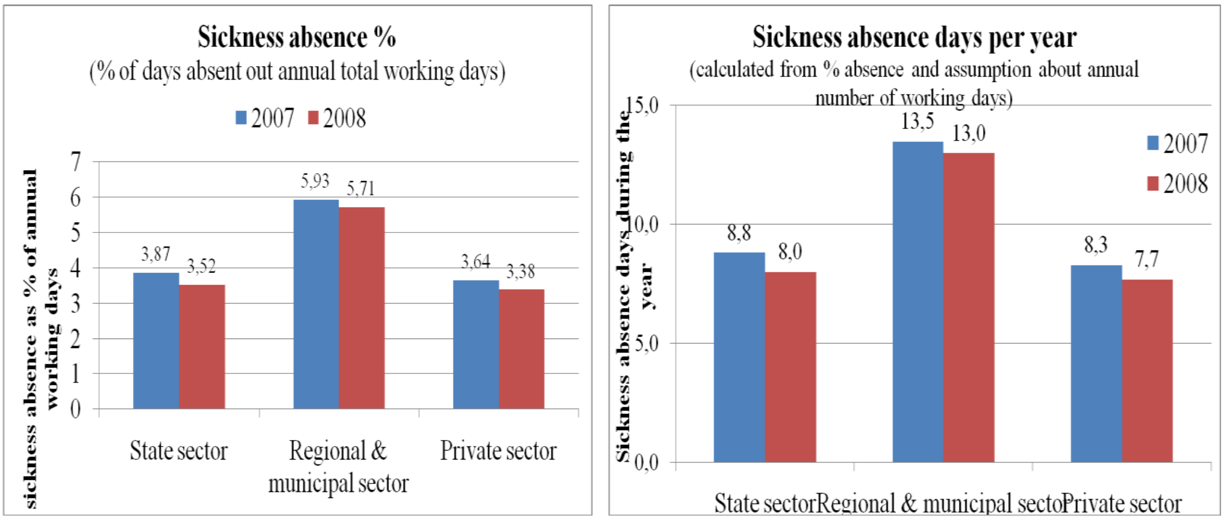
Many working days are lost due to sickness absence. The official Danish statistics are shown in figure 1 based on employer-reported absence information. According to these numbers more than 4% of the total number of annual working days is lost in this way – with considerable variation across sectors of the economy. Measured in absolute number of days the average across the sectors is between 9.5 to 10.2 days per employee[6].

For long term sickness absence the public sector pays compensation to employers. For most occupational groups compensation is a relatively small fraction of the actual wages. 'Long term absence' is defined in the relevant legislation[VII]. As of April 2007 the period was changed from 14 to 15 days of absence, and as of July 2nd 2008 the period was extended to.21 days. This means that for the first 14 (15) days and from mid 2008 the first 21 days of absence, including week-ends, the employer pays for sickness absence (essentially full pay). After this period the employer receives compensation from the public sector

[VI] In the (epidemiological) literature there is no established definition of 'long term' absence. In the present context the term normally refers to the definition used in the legislation underlying sickness absence compensation.

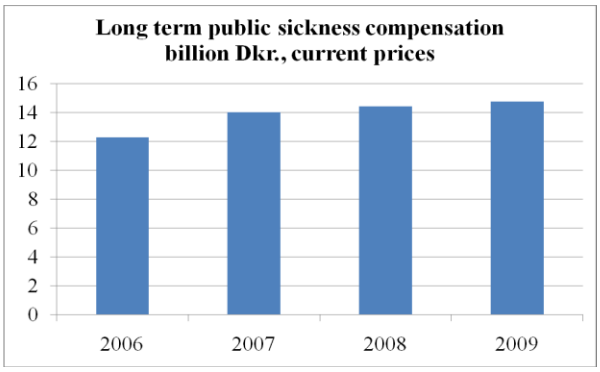
[VII] The act on sickness compensation (Sygedagpengeloven, lov nr 563 af 09/06/2006 with subsequent changes). See Johansen et al[7] for legislative changes and sickness absence philosophy in Denmark since 1973.

Figure 1: Percentage and absolute number of annual working days lost due to sickness absence.



Source: Statistics Denmark[6].

Figure 2: Amount of public sickness compensation, billion Dkr. 1 € = 7.50 Dkr.; Source:[6]



(administered by the municipalities). Most often the employer tops this compensation so that the sick listed employee receive full pay. However, the rules and practices concerning this vary by union contracting domain and by company. There is no calculation available showing the total costs of short and long term sickness absence, but the public costs of sickness compensation are shown in figure 2.

The Danish absence percentage, cf. figure 1, is relatively low compared to other countries[8], despite what internationally may be considered high Danish compensations rates, i.e. essentially full pay for at least the first 21 days and often also after this period.

Over the past decade there has been increasing interest in trying to decrease sickness absence, and in particular long term absence[9-13]. There are several reasons for this. First and foremost, with a decreasing workforce – and until late 2008 a record low unemployment rate – one way of increasing the number of working days available in the economy is to decrease sickness absence. As an example: the long term sick listed at any one time make up 7-8% of the working force, and if converted to full time equivalents, FTEs, it is around 90,000 FTEs – which in 2007 and the first part of the 2008 was more than the number of unemployed[11]. Thirdly, it turns out that being long term sick listed increases the risk of deroute in the sense it may be the beginning of the path towards disability pension. For instance, only 25% of persons who have been sick listed for more than one year return to work, while 90% of those who have been sick listed for less than six months return to work[11].

It is increasingly realized that active and early follow-up of long term sick listed employees is important in order to avoid not only keeping them in the sick role but also to prepare them – if possible – for return to active work. The question, however, is what type of intervention is relevant and needed? There is emerging evidence that well-coordinated and timely support from health and social services is important and may shorten the period of sickness absence[11, 14, 15].

A Swedish report on sickness absence noted that a crucial factor in early/earlier return to work was better and efficient cooperation between primary health care and social care.[16].

OECD recently published a synthesis of findings across OECD countries, including Denmark, and noted that “in particular, it is essential to better direct the actions of general practitioners by emphasizing the value and possibility of work at an early stage, and then to keep the sickness absence period as short as possible …”[17]

In other words there is some, but in no way overwhelming or convincing evidence that timely/fast access to and use of health care is important in order to decrease (long term) sickness absence. Or put negatively: Unnecessary waiting time for treatment may be a barrier to early return.

It seems intuitively correct that sickness absence not only in many cases leads to utilization of health services, but that use of services also most likely ought to shorten the period of absence compared to, ceteris paribus, identical persons not using health services or use of services with some delay (waiting time). However, whether it happens simultaneously or time-lagged is unclear. For this paper it has only been possible to identify three studies looking at this rather obvious relationship between use of health services and sickness absence in the rather voluminous literature on sickness absence[18-20], and of which only one[20], not yet published in a scientific journal, is directly relevant here. In that study it was concluded that almost all waiting for health care had a statistically significant impact on the duration of sick leave. However, there is no available evidence on which type of health care is the most relevant, e.g. consultation with an occupational physician, GP, or physiotherapy. Among other things this obviously depends on the nature of the illness underlying the sickness absence.

In a Danish context – but not internationally – there has been discussion of the effect of health insurance on sickness absence – triggered by the issues mentioned in the introductory section and in particular the separate issue of justifying the tax exemption by trying to document public savings on sickness compensation to long-term sick listed[VIII].

At the outset it is obvious that health insurance per se does not influence the length of sickness absence. Rather, at best health insurance is an ‘enabler’ – possibly enabling faster access to (private) health care than is the case for non-holders of health insurance. This raises a double issue: is health insurance put to actual use in case of sickness absence, and is the privately provided health care received more timely and better coordinated than health care used by non-insured?

Two analyses have addressed the relationship between health insurance and (long) term sickness absence[4, 21]. DSI found no difference between health insurance holders and non-holders regarding sickness absence based on the 2005 version of the SUSY-survey (national survey of illness, absence, health status, health behavior etc.) whereas Borchsenius and Hansen based on register data on compensation for long term absence linked with insurance data using propensity score and matching estimators found a significant and considerable decreasing effect on long-term absence for insurance holders compared to non-holders.

However, in none of the analyses was the logical question of why having health insurance per se should influence sickness absence addressed. It seems quite clear, as noted earlier, that the real underlying issue must be to what extent the insurance has been used to gain access to and use of health services and whether this use was linked to a spell of sickness absence.

However, not only are there other ‘interventions’ than health insurance and/or health care available to decrease short or long-term sickness absence, e.g. stress management[22], cognitive therapy[23, 24] or active involvement of the employer[25, 26], but there is also a host of other determinants of (long term) sickness absence than access to and use of health care services, e.g. a social gradient, work and environment – and at least for short term absence - most likely more important than health care. A considerable Danish literature on risk factors for long-term sickness absence has been published over the past decade[27-36]. Similarly there are a few works on the effect of health behavior, e.g. exercise, smoking, and alcohol consumption on long term sickness absence[37] or work place design, e.g. changing work environment, to prevent sickness absence This literature is relevant in the sense that if one wants to isolate the effect of use of health services in general or use of specific health services on length of sickness absence, one needs to control statistically for work environment, social gradient variables, and the like that may jointly determine both sickness absence and health service use.

.

[VIII] Whether this is a valid and general argument in favor of tax exemption is a separate issue not discussed here.

Theoretical framework

In the epidemiological literature there is an amazing lack of a theoretical framework for understanding and analyzing sickness absence. For an exception see Labriola[38]. Much of the literature must be classified as exploratory building on and consolidating earlier results. Some empirical clear results are emerging, however, i.e. the effect of work environment, type of work, and social gradient variables are important. The situation in the economic literature is somewhat better concerning theoretical framework.

In the first review of the economics of absence[39] from 1996 Brown and Sessions noted that the area was underdeveloped relative to other areas of labor economics. They went on and noted that in the models of absenteeism based on the traditional static neoclassical labor supply theory (work – leisure choice) absenteeism essentially was based on the premise that it arises not because the individual is unable to work, but because he/she chooses not to, i.e. absence is voluntary and due to an attempt to adjust, if possible, to a utility maximizing position[IX]. It is a striking weakness as of 1996 that theoretical models of labor supply ignored health status of the individual[X] - and for that matter other determinants of sickness absence. Empirical works by economists is not always based on an explicit theory or, if the case, standard labor supply theory, e.g. Allen's 1981 classic[40]. The model does not include health status[XI]. In Allen's empirical work, he, however, included indicators of health/ill health or indicators of harmful health effect, e.g. 'dangerous work place'.

Over the past 10 years much progress has been made in the economics of absence. As is to be expected much of the literature focuses on the effect of economic incentives, i.e. either within an efficiency wage[XII] framework or focusing on the payment/remunerations structure and/or degree of compensation in case of sickness absence – and hence within the traditional choice framework – disregarding for instance accidents at work and the like (involuntary absence): "The analysis of

[IX] Allen illustrates this clearly: "When a worker contracts for more than his desired hours given w, he retains an incentive to consume more leisure. One way of doing this is to be absent from work." (p. 78)[40]. Economists are amazingly naïve – with greater faith in models than 'real world' observation.
[X] The earliest exception is probably Barmby[41] who in an attempt to move away from the supply-orientation introduced employer monitoring of effort, and hence absenteeism/shirking. To this end he introduced asymmetric information regarding the health status of the employee.

[XI] In a 1985 working paper[70] we developed a model based on Becker's allocation of time model[48] and used an elaborate set of health status variables in the empirical estimation of the model – at the time, the most exhaustive set available anywhere - and found that the inclusion of health status very much influenced the estimation results.

[XII] For the sake of clarity, following the New Palgrave Dictionary of Economics 'efficiency wages' is a term used to express the idea that labor costs can be described in terms of efficiency units of labor rather than in terms of hours worked, and that wages affect the performance of workers. The incentive effects of wages stem from the effect of the level of compensation on the cost to the worker of being fired. Thus, wages above the market clearing level will increase effort, decrease employee theft, decrease absenteeism, and decrease quits. – The classic article is the 1984 shirking model by Shapiro and Stiglitz[42] where the problem is posed in terms of moral hazard. In these models absence is supposed to reveal the employee's level of effort.

sickness absence is placed firmly in the agenda of economics by the idea that sickness absence rates are the consequence of choices that can be mediated by economic (and other) incentives."[43]

The theoretical models can be grouped into three main (somewhat overlapping) groups following Chatterji and Tilley[44]. 1 The supply side approach, 2. The efficiency wage approach and 3. The contract approach. The latest addition is a model type based on the health capital/production of health.

The labor supply approach has already been outlined above. The main point is that sickness absence modeled within the work-leisure framework is a choice variable, in part due to working hours being fixed exogenously, e.g. through union contracts. If more leisure is desired this is done through sickness absence meaning that absence is shirking and not rooted in underlying health problems or accidents at work. As noted, Allen was one of the first to use this modeling approach[40].

Barmy et al[41] were among the first to recognize that employees may be absent with or without good cause. They used the efficiency wage approach, see footnote XI, whereby, among other things, the actions of the employer could be modeled. A more recent example of the wage efficiency approach is the work by Ose[45]. In her model she tries to separate the effects of voluntary absence and absence related to ill health, where health effects are assumed to be tied to working conditions At the general level her modes builds on and extends the classic 1984 Shapiro and Stiglitz efficiency wage model[42].

The contracting approach goes back to Coles and Treble[46] who looked at the issue from the employer perspective. Workers can be either absent with cause, choose to be absent without cause or choose to be at work. The employer can only observe the absence-attendance choice of the employee. The challenge for the firm is to choose some wage-sick pay contract so as to maximize profit subject to a zero profit condition and an incentive compatibility constraint.

Like in the other two approaches the focus is essentially on economic incentives and asymmetric information. Other causes of absence are not really included, e.g. the working environment (physical and physic).

Turning to the models used in health economics, in particular the tradition developed by Grossman (se next section), Afsa and Givord have developed a model with explicit inclusion of health status and working conditions[47]. This is the first example of a possibly new class of models. .

What is important is the need for inclusion of health status and of working conditions. This is done in the following model, to a certain extent also using Grossman's approach.

A simple model of the 'demand' for sickness absence

In his theory of the allocation of time Gary Becker[48] outlined a model where households are seen as producers of commodities instead of solely consumers of goods and services. Grossmann in his path breaking work on the demand for health[49] used Becker's basic idea of household production and turned it into a 'health production approach'. He defined health as a durable capital stock that implies that the end product is not health as such but the services flowing from this capital good. In Grossman's formulation, individuals derive utility from the services that health capital yields and from the consumption of other commodities. The stock of health capital depreciates over time, and the consumer can produce gross investments in it according to a household production function using medical care and their own time as inputs. It is assumed that the efficiency of the production process depends on individuals' stocks of other forms of human capital, especially education. The return from the stock of health capital may be defined as the total number of healthy days in each year, which generates utility directly, since being healthy yields utility (termed the "consumption" motive in the literature), and indirectly, since being healthy yields income which in turn can be used to purchase goods or to produce commodities which influence utility (termed "investment" motive in the literature).

For readers not familiar with the Grossmann model, the main ideas are depicted in figure 3 below taken from an early Danish study[50, 51]

Inputs		Capital stock		Flow of services (outcome)
Health care				
Drugs				
Genetic endowment	=>	stock of health	=>	healthy time
Education				
Lifestyle				
Work environment				
Housing				
...				

Figure 3: The basic idea behind the health production function

Without elaborating further is should be clear that at the general level a good point of departure for a model would be a health production function[XIII].

$$(1) \qquad h=h(\boldsymbol{\sigma},\mathbf{q},\mathbf{X})$$

h is health status, e.g. self assessed health, which take on high values for good health. The vector **σ** describes 1…n possible health shocks like onset of a disease, worsening of a chronic condition, or accidents that the individual has experienced. The vector **q** expresses experienced access to health care, for instance waiting time, (private) health insurance. Lastly, **X** is a vector of personal and job characteristics like education and work environment. Of course one could have separate vectors for personal and job characteristics. The important point is that health is influenced by, among other things, both individual and work place aspects.

[XIII] The following model is essentially equivalent to the one presented by Granlund[20] but with different interpretations and explanations.

The utility function is as follows

$$(2) \quad U = U(h,b,z, \mathbf{X})$$

where the b is consumption of market goods and z is leisure time. The utility function is characterized by $\delta U/\delta b > 0$; $\delta U/\delta z > 0$; $\delta U/\delta h > 0$; $\delta U^2/\delta z \delta h < 0$; $\delta U^2/\delta^2 b < 0$; $\delta U^2/\delta^{2z} < 0$; $\delta U^2/\delta b \delta h < 0$; $\delta U^2/\delta b \delta z < 0$.

By normalizing the time endowment to unity, leisure can be defines as $z=1 - l + a$, where l is the number of scheduled/contracted working hour and a is sickness absence. This implies, however, that absence is considered on par with leisure, i.e. no 'disutility'. Alternatively, but not done here, one could distinguish between disutility of absence and disutility of work effort.

With these preliminaries the budget constraint can be defined as

$$(3) \quad wl + y - (1 - \beta)wa = b$$

with w being the wage rate, y non-labor income and β is the share of the wage the employee receives when absent ('compensation rate') .and the price of consumption goods is normalized to one. It is implicitly assumed that health care is free.

By substituting for h, b, and z in the utility function, (2), using (1), (3) and the time constraint, the first order conditions for worker absence can be written

$$(4) \quad \delta U/\delta a = \delta U/\delta b \, (1-\beta)w + \delta U/\delta z = 0$$

In general $\delta U/\delta b$ and $\delta U/\delta z$ may also depend on, besides a, w, l, y $\mathbf{q}$, X and $\boldsymbol{\sigma}$**.** Hence, the 'demand function' for sick absence can be written as

$$(5) \quad a = a(\mu, c, l, \boldsymbol{\sigma}, \mathbf{q}, \mathbf{X})$$

The means that sickness absence is a function of the employee's potential income ($\mu = wl + y$)), the cost of absence ($c=(l - \beta)w$)), health shocks, access to health care/waiting time, **and** various individual and job characteristics.

Equation (5) thus provides (at least some) justification for the regression analyses reported in table 7 and table B in appendix I.

In order to show how 'demand' for sickness absence depends on μ, c, and l and one or more of the vector-elements of access to care (e.g. health insurance) and waiting for care, e.g. q_1 in $\mathbf{q}$, one can differentiate eq (4), using the implicit function theorem, with respect to a and one of these variables one at a time, thus generating hypothesizes about expected sign, e.g da/dq_1 where q_1 might be health insurance (waiting time). This line is not pursued here. However, it should be noted that the imply-cation of da/dq_1 is that waiting time, by its negative effect on health , increases the demand for absence – and leaving out intermediate mediating effects, see p. 9 in Granlund[20] – meaning that prolonged waiting time (or patchy coordination of services etc) increases the 'demand' for health and by implication the duration of sickness absence.

Eq. (5) is the point of departure for the quantile regression model below. However, health insurance has not been included in the above model. The following section addresses this, but in a less forma way. This section should be seen in connection with the empirical work on the propensity score.

Health insurance: ex ante and ex post moral hazard

There exists a 100% US focused literature on health insurance and the labor market[52, 53], largely due to dominance of employer paid health insurance in the US[XIV]. It is, however, almost irrelevant in the present context. In part because it concerns full health insurance, i.e. both for acute and elective care, in part because it is empirical and largely atheoretical[XV]. In addition it must be seen in a US institutional context. In the present context models of firms choosing to pay for health insurance for their employees ('firms' demand for health insurance') would be needed, but only two (US) articles on employer decision models for health insurance have been identified[54, 55] . Therefore, as the aim is not theory development, the following is some general observations stemming from the general insurance literature.

In health economics a key question is whether or not complementary/duplicate health insurance encourages moral hazard in the use of health care, i.e. in 'excess' of the level of use without health insurance. Moral hazard occurs when the behavior of the insured party changes since the insured party no longer bears all or just some of the costs of that behavior. In consequence the insured have an added incentive to ask for pricier and/or more elaborate medical service, e.g. timely without (too long) waiting time. In these instances, individuals have an incentive to "over consume".

Having health insurance may induce two types of behavioral change – at least according to the conventional wisdom. One type is the risky behavior itself, resulting in what is commonly called *ex ante* moral hazard. In this case, insured parties behave in a more risky manner, i.e. health promotion and preventive activities may be neglected – privately or at work[XVI].

[XIV] Superficially there are similarities to the Danish situation for employer paid health insurance, namely that the employer paid premium is not treated as taxable income to employees – and that employee payment for insurance is tax deductible as well (a certain similarity to the Danish gross-deduction arrangement ('brutto-træksordningen'))

[XV] Gruber notes as of 2000: "…the previous point reflects the atheoretical nature of this literature. While the empirical innovations in this area have been impressive, the theoretical have been much more modest. If this literature is to move beyond its infancy …a firmer theoretical underpinning will be necessary"[53], p. 700-701.

[XVI] Insurance companies try to counteract this by offering lower premiums if for instance a work place has health promotion and preventive activities for employees in place. This is the trend in the Danish health insurance market. One of the reasons to include 'health promotion at the place of work' in the quantile regressions below can be found in the issue of ex ante moral hazard. A number of health behavior variables are available in the health insurance (HIS) data set, but have not been put the use.

The second type of behavioral change is the reaction to the negative consequences of risk once they have occurred, i.e. people have fallen ill and/or are absent from work, *and* once insurance is provided to cover totally or partially health care costs and/or insurance gives access[XVII] to alternative sources of medical care, i.e. in the private sector. This leads (again according to conventional wisdom) to a total level of health care consumption that is higher than in a world without health insurance. This is often called *ex post* moral hazard. For example, without (employer) paid health insurance, most persons have to rely on free medical care provided by the publicly financed health care system, possibly with waiting time and/or patchy coordination of care. *Ex post* moral hazard in the present context then concerns two overlapping issues: increased consumption of medical care and access to alternative sources of care in the private sector (at least in the case of Denmark).

Health insurance is only indirectly linked to sickness absence. As noted above, health insurance per se cannot be assumed to affect sickness absence[XVIII] - with the possible exception of the situation hinted at in footnote XIV – but even then it requires actual use of services to have an effect. The effect of health insurance must be indirect: sickness absence (may) lead to treatment of the underlying illness, and health insurance may then facilitate medical care provided outside the public health care system, most likely at private hospitals. This can either substitute publicly provided health care or supplement it.

As health insurance and sickness absence it should be noted that the issue of ex post moral hazard has been and still is being analyzed using the HIS data set[XIX]. *Preliminary* results[5, 58, 59] from the HIS data seem to indicate that moral hazard seems to be negligible or absent with the possible exception of physiotherapy. The relevance of this in the context of matching estimators of effect of health insurance presented below is that the effect of health insurance on sickness absence probably is not be explained by 'over-consumption' per se. Unfortunately the data does not enable us to decide whether private health care is a substitute for publicly provided health care (however, see table 5 and comments in connection with the table – using data from the presenteeism data (PRS data set) or that the private health care may be more accessible than public health care, i.e. less waiting time, see table 6A. Whether private health care is better coordinated without (unnecessary) time delays unfortunately is not described in either of the data sets used here.

XVII Conventionally it is assumed that taking out insurance is rooted in risk aversion. However, **access** to otherwise too costly services for the individual, e.g. treatment at private hospitals, may be another and maybe stronger reason to take out health insurance. Nymann has argued this point[56, 57].

XVIII In 'causal terminology': Health insurance cannot cause a change in sickness absence. At best it can facilitate change. However, to be of policy value the causal mechanism must be understood, i.e. the (causal) effect of consumption of private medical care.

XIX Astrid Kiil in her upcoming ph.d-thesis (late summer 2011) addresses this in two ways. In chapter 6 the theme is "Does employment-based private health insurance increase the use of covered health care services? A matching estimator approach" and in chapter 9 where the issue addressed is "An empirical comparison of methods to identify the effect of voluntary private health insurance on the use of health care services"

Data

Two survey data sets are available. The first, '*the health insurance survey'*[11]*, (HIS)*, is a cross sectional survey from June 2009 of the Danish population aged 18-75. It is fairly representative of the population in this age bracket. The sample size is 5,447 respondents of which 3, 470 were occupationally active. The present study focuses on the latter. The individuals in the sample answered an extensive internet-administered questionnaire focusing on voluntary health insurance, risk aversion, socio-economic variables, use of health services, and also a question about sickness absence from work the past twelve months. In this survey and the following on presenteeism sickness absence in consequence is self reported for a period of 12 months. It is a key variable in the empirical analysis of the effect of health insurance. The literature does not give much guidance on the optimal reporting period or the accuracy of self report absence[60] compared to other sources (that may also contain bias). The self-reported insurance status is another important variable. When the numbers reported in the next section are compared to publicly available data there is no reason to believe misreporting is of great importance.

The second data set, *the presenteeism survey (PRS)*, is also a cross sectional survey, but of the occupationally active population only. It was carried out in December 2010. The sample size is 4,060. Respondents answered an internet administered questionnaire aimed at presenteeism ('sick at work'), absenteeism with a clearer distinction between short and long term absence than in HIS, work conditions, health insurance and the use of health services. Some of the questions were aimed at in some detail to try to understand the use of health insurance to obtain health services. Several of the questions are identical to the ones used in HIS.

Both surveys were preceded by pilot testing, N>100. Some of the questions were identical in the two surveys, e.g. questions about insurance and 12 months sickness absence.

Descriptive results

Coverage with health insurance and use of insurance

In both surveys there is information on the following types of insurance[XX]:

- employer paid health insurance for employees
- coverage through spouse' s employer paid health insurance
- privately paid health insurance in commercial insurance companies
- privately paid sickness insurance[XXI] taken out through the non-profit company 'denmark'

[XX] The first three bullet points were preceded by the following text in the questionnaire: "An increasing number of companies offer their employees health insurance. A health insurance covers expenses to operations at private hospitals among other things, and usually also counseling and treatment by physiotherapists and chiropractors. The main rule is that the employer pays the insurance premium". Hence, a clear definition/delineation has been provided.

There seems to be good agreement on important insurance questions across the two surveys. In both surveys 37-38% of the respondents indicated that they were covered by an employer paid health insurance. Between 7-8% answered that they were covered via the spouse's health insurance. Regarding private health insurance the difference between the two surveys was wider: 7% (HIS) and 10% (PRS). This is slightly more than sample uncertainty can explain. However, in PRS a filter question was used. Also, it should be remembered that the two surveys were carried out 18 months apart. Hence, the private health insurance market may have grown.

In both surveys respondents were also asked whether they had used the health insurance to gain access to (private) health care within the past 12 months. In the HIS-survey 21% of insurance holders had made use of it within the past 12 month while it was 25% in the PRS-survey. In addition to sample variation the difference may also be caused by the reported increasing use of health insurance over the almost 1,5 year separating the two surveys.

Some descriptive results for health insurance and sickness absence

To give a 'feel' of the core issue in this paper some descriptive data are presented below.

In a simple univariate context there is marked difference between days of sickness absence and insurance status in both surveys, table 1, possibly mirroring possible selection effects at the company level because it is the companies that decide to offer employer health insurance to their employees. In view of the generally declining sickness absence over the past two years the differences reported in table 1 to some extent is understandable. In the HIS data 32% had no days sickness absence, while the corresponding number for PRS is 33%.

Table 1: Summary of days of sickness absence the past 12 months according to insurance status

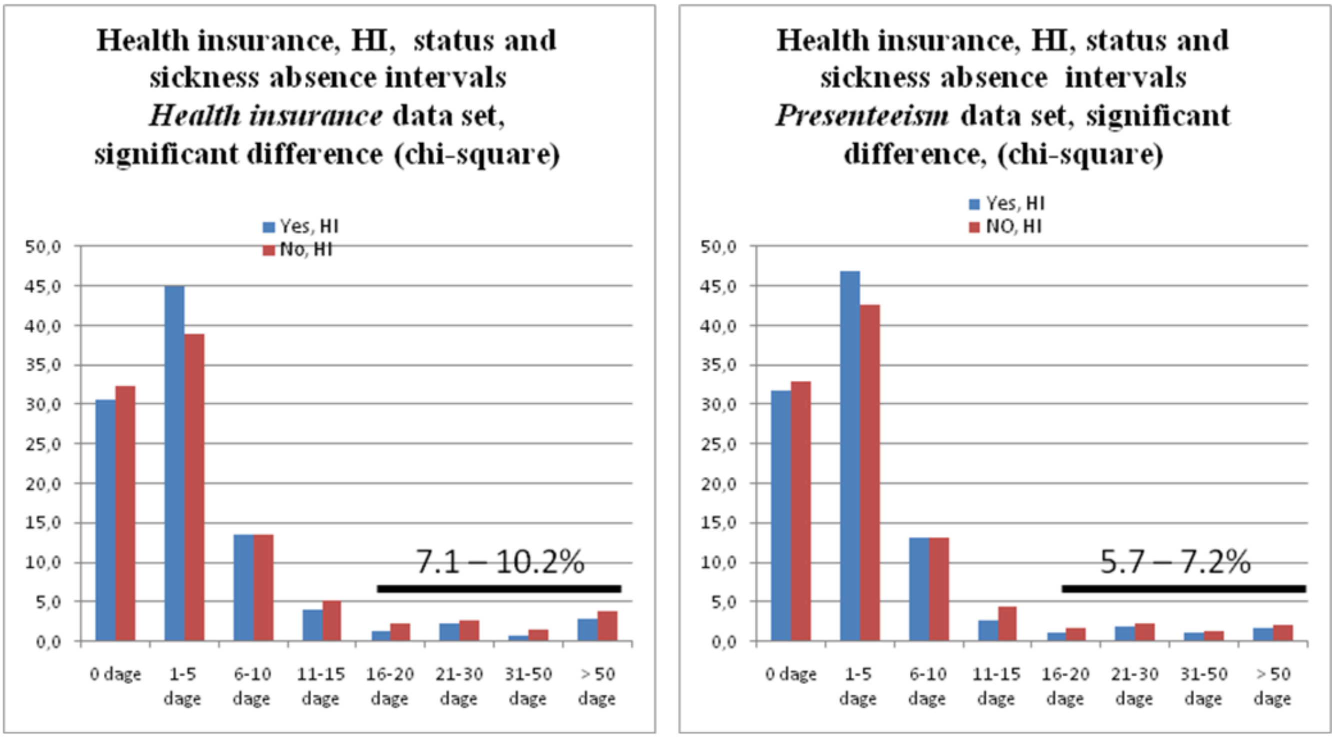
	Health insurance survey, HIS: Days of absence	**Presenteeism survey, PRS: Days of absence**
Yes, has health insurance	7.1	5.8
No, do not have health insurance	9.4	6.4
Don't know	12.7	7.2

Long term illness according to the act on sickness absence compensation means being absent for three consecutive weeks, or 15 workings days disregarding week-ends. In the HIS data 10% had more than 15 days of absence, see figure 4 for details.. However, from the survey it cannot be determined whether this concerns consecutive days. In PRS data the percentage was 8%. In PRS it

[xxi] Throughout this paper the term 'sickness insurance' is used for insurance taken out through 'denmark', while the term health insurance is used for employer paid insurance, inclusive of coverage via spouse's health insurance. 'Private health insurance' refers to privately paid health insurance taken out via commercial insurance companies.

is possible to distinguish the group with 15 or more consecutive days of absence. 6% of the sample had had more than 15 days of consecutive days of absence, i.e. long term sickness absence as defined previously. 61% of the group with 15 or consecutive days of absence did not hold health insurance, HI.

Figure 4: Sickness absence and health insurance (HI) in the two datasets (% along the y-axis).



Without being causal table 2 shows that those who made use of the health insurance also had higher sickness absence than those who did not – and markedly so. This can come as no surprise: they use health insurance because they are sick!

Table 2: Use of health insurance and sickness absence during the past 12 months

	Health insurance survey, HIS: Days of absence	Presenteeism survey, PRS: Days of absence
Made use of health insurance within the past 12 months	12.8	10.0
Have not made use of health insurance within past 12 m.	6.2	4.3

As noted above the Presenteeism survey, PRS, included several questions designed to throw light on the role of HI and sickness absence not included in HIS (out of which PRS grew 18 months later). In the PRS a question directly addressed the question of using health insurance to gain access to health care in connection with sickness absence. 30% (120) of those who had made use their health insurance in the past 12 months confirmed that it was related to sickness absence. Put differently, of the 1537 respondents who indicated to have HI, 8% use it of in connection with their short-or long(er) term sickness absence, and 3% only used it solely for this purpose. In other words, if the effect on sickness absence is to result in a change of the duration of the average sickness absence period, the effect for the active user group has to be substantial[XXII]. In addition the same group also has had access to publicly provided health services.

Table 3: Relationship between use of health insurance and sickness absence (PRS only)

You have used your health insurance within the past 12 months. Was the reason your sickness absence?	**Mean number of days absent**	**N**
Yes, the only reason	34.4	40
Yes, partly the reason	12.0	80
No	6.1	281

Furthermore, in the PRS there was a question of more general nature, namely whether it was the respondent's experience or impression that having health insurance calls for quicker access to service, for instance shorter waiting time/more speedy booking of consultation. 60% of the HI users answered positively to this question, table 6A.

Table 4 shows the use of private hospitals and sickness absence. It is striking to note the number of sickness absence days for HI-holders undergoing surgery.

Table 4: Operations and/or MR, CT scans and X-ray at private hospitals (PRS dataset)

	Mean number of days of absence	N
Operation at private hospital Yes (do have HI and have been operated 1 or several times past 12 months)	19.8	52
No (do have HI, but not operated)	8.7	368
Do not have HI or have not used HI	5.9	3,540
MR, CT, X-ray at private hospital		
Yes	21.2	79
No	7.5	341
Do not have HI or have not used HI	5.8	3640

XXII A quick calculation illustrates the issue: The average number of days of sickness absence for all HI holders is 5.8 days (table 1). If one were to assume that the absence days in table 3 for those who confirmed that they used the their HI totally or partially in connection with sickness absence were reduced to 0, then the average for the total HI-group would be reduced to 4.2 days or 4.9 days if the 40 who only used HI for sickness absence purposes were included in the calculation.

However, health care from the public sector health care of course may also be relevant for sickness absence. Table 5 shows hospitalization and sickness absence

Table 5: Hospitalization and sickness absence (PRS)

	Days of sickness absence	N
Has not been hospitalized	4,72	3731
Has been hospitalized	23,48	329

Of the 52 persons in table 4 who had had an operation at a private hospital 13 (25%) had also been hospitalized at a public hospital, i.e. there is not perfect substitution and the use of health services cannot without problems be compartmentalized to 'private' or 'public'.

Two observations emerge from table 6. First, health insured have a better self rated health status compared to non-insured or don't knows. This is reinforced by the fact that health insured have an average of 0.73 chronic illnesses (out of 14 possible) compared to the group without health

Table 6: Health status, insurance status and sickness absence (PRS data-set)

Health status **Has HI**	Really good	Good	Passing	Bad or really bad	Total
Days absent	2.5	4.3	10.5	23.1	5.8
%	17.7	59.1	20.2	3.1	N=1,537
Do not hold HI					
Days absent	2.9	4.6	9.7	20.4	6.5
%	15.7	53.5	25.5	5.3	N=2,265
Don't know					
Days absent	2.9	5.7	8.5	30.8	7.2
%	15.3	58.5	21.0	5.2	N=248

insurance that on average had 0,9 chronic illnesses. Secondly, sickness absence clearly varies by health status.

There may be an access motive in holding HI (see section on ex ante and ex post moral hazard). This should give faster access to health care in the private sector compared to the public sector. In the PSR data a specific question address this, table 6A. The question asked was directed at those who had made use of their insurance within the past 12 months: "According to your experience does holding health insurance mean that you gain faster access to diagnostic procedures and treatment (quicker clarification of your situation), compared to not holding health insurance?"

Table 6A: Faster access to health care for HI holders (PRS data set)

	N	Percent
Yes	248	60.05
No	86	20.82
Don't know	79	19.13
Total	413	100.00

Quantile regression

The statistical analysis will proceed in two stages. In stage 1 the analysis strategy is quantile regression analysis with number of sickness absence days as dependent variable. In stage 2 propensity score and matching estimators will be used.

In the sickness absence literature there is a long tradition for analyzing the determinants of sickness by regressions analysis/logistic analysis – and often the sample is split into two: short and long term absence (however defined). However, by using quantile regression the latter is avoided, far better use of all observations is made, and one can study how determinants may change across the conditional quantiles. This is a new approach in the existing sickness absence literature. However, only under rather restrictive conditions can one interpret the coefficient of health insurance – one of the 'determinants' of sickness absence – in a causal sense. Therefore, an analysis of treatment effect using propensity score matching is carried out that, at least in principle, allows a (more) causal interpretation of HI.

There is no doubt an important difference between short and long term sickness absence. The underlying 'causal' mechanisms differ in important ways, e.g. underlying illness, outright shirking or a flu versus pneumonia or cancer – or long term consequences of an accident. It would be tempting; therefore, to split the sample into two, one for short term and one for long term illness. However, a better strategy is use quantile regression[XXIII] to study (conditional) differences across the sample[XXIV]. Koenker and Hallock explain this clearly: "We have occasionally encountered the faulty notion that something like quantile regression could be achieved by segmenting the response variable into subsets according to its unconditional distribution and then doing least squares fitting on these subsets. … Clearly, this form of truncation on the dependent variable" would yield disastrous results in the present example. In general, such strategies are doomed to failure for all the reasons so carefully laid out in Heckman (1979*). It is thus worth emphasizing that even for the extreme quantiles all the sample observations are actively in play in the process of quantile regression fitting.*"[62] (italics added).

Median regression, a special case of quantile regression, estimates the median of the dependent variable, conditional on the values of the independent variable. This is similar to least-squares regression, which, however, estimates the mean of the dependent variable. Median regression finds the regression plane that minimizes the sum of the absolute residuals rather than the sum of the squared residuals. Moving on, the quartiles divide the population into four segments with equal proportions of the reference population in each segment. The quintiles divide the population into 5 parts; the deciles into 10 parts etc. The quantiles, or percentiles, or occasionally fractiles, refer to the general case. In quantile regression these ideas are extended to the estimation of conditional

[XXIII] This is not the place for even a brief exposition. See Koenker for a full exposition[61] and Koenker and Hallock for an excellent easy to follow exposition[62].

[XXIV] Quantile regression also allows the estimation of treatment effects using instrument variables[61]. See chapter 7 in Angrist and Pischke[63].

quantile functions - models in which quantiles of the conditional distribution of the response variable are expressed as functions of observed covariates. By using quantile regression it possible to study different parts of the conditional distribution of sickness absence, e.g. looking at the 90 og 95% part to the distribution it is possible to look at long term sickness absence. The use of quantile regression is far better suited to study sickness absence the conditional mean approach of conventional OLS, in particular if long term absence is of interest. .

The coefficients in quantile regression models are interpreted in the same way as in ordinary OLS regression models, with the caveat that they are partial effects on the respective quantile as opposed to the mean for OLS.

The dependent variable in the following is sickness absence within the past twelve months. Independent variables have been chosen based on both the theoretical model above (eq. 5) and the epidemiological exploratory analyses referenced in the background section.

1. Socio-economic variables
    - Personal_income
    - Age
    - Gender
    - Children
    - Education
    - Union membership *(HRS data set only)*
    - Management responsiblity
        i. *(only for PRS-dataset)*
2. Health variables
    - Self assessed health status
    - Number of chronic diseases,
    - Long term illness, long term consequences after accident etc.
3. Use of health services
    - general and specialist practice visits
    - A&E visits and outpatient visits, including same day surgery
    - Hospitalization
4. place of work
    - Public-private,
    - Number of employees
    - Health promotion activities at place of work,
    - Work place policy on sickness absence (*only for PRS-dataset)*
    - Sickness interview at place of work (*only for PRS-dataset*)
    - Physically demanding work *only for PRS-dataset)*
    - Physically exhausted (*only for PRS-dataset)*
    - If absent, then others take over my tasks (*only for PRS-dataset)*
    - Overall satisfaction with place of work (*only for PRS-dataset)*
5. Health insurance variable
    - Holds employer paid insurance
    - Use of health insurance in the past 12 month
        i. information available on specific services received, including whether triggered by sickness absence.

Table 7[XXV] shows, for comparison, the result of an OLS, then the results from the quantile regressions (median (q50), q25, q90, and q95) and the last column shows the regression of difference in quantiles, here between q25 and q95 (interquantile regression)[XXVI]. In essence one can consider q90 and q95 as analyses of long term sickness absence.

The two main variables of interest here: 'Health insurance' and 'use of health insurance within the past twelve months' are shown at the bottom of the table.

The health insurance coefficients ('not having HI' compared to those having) are not significant[XXVII], i.e. do not exert a significant influence on sickness absence – be it short or long term (compare q25 and q95) and the interquantile coefficient in the right hand column of the table confirms this. Furthermore, compared to those with health insurance, those without had lower sickness absence regardless of which quantile that is considered.

The 'use of health insurance' variable also does not show any significant effect on sickness absence[XXVIII]. Looking at these coefficients[XXIX], it is seen that those who had not used their health insurance had fewer days of sickness absence compared to those who had. A similar observation holds for the group without health insurance. This really should come as no surprise, because one must assume that those who made use of their health insurance had more serious underlying (medical) problems than those who either did not use the insurance or those without insurance.

[XXV] To retain as many observations as possible 'don't know' has been retained as separate categories. One may argue that a better alternative would be to impute values.

[XXVI] To make comparison easier the standard errors and confidence levels are not reported. One may then directly – column by columns – compare coefficient estimates. SE and CI are, of course, available on request.

[XXVII] In the STATA manual on qreg it is indicated that the standard errors are sensitive to the number of bootstrapping replications. Some experimentation, i.e. 100 vs. the standard 20, seem to indicate that the obtainde results are fairly robust in this regard

[XXVIII] To conserve space the two coefficients – from a separate regression analysis – have been inserted in table 7 that contains the coefficients from the regression where 'having HI' was estimated

[XXIX] Here inserted as part of table 7 with the coefficients from the full analysis with 'health insurance' as dependent variable. The coefficients for 'use of health insurance' were estimated in separate analyses, but inserted in table 7 for easy reference. The other coefficients in the analysis of 'use of health insurance- were note markedly different from those shown in tables 7.

Table 7: Regression results (OLS, quantile), PRS-dataset: (STATA, sqreg, iqreg 20 rep. for bootstrap). Significance levels indicated by stars, see below table

Dependent: days of absence past 12 months	**Ord OLS**	**Median (quantile)**	**q25 (quantile)**	**q90**	**q95**	**q25-q95 quantile difference-**
Characteristics of the employees	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
Health						
Number of chronic diseases	0,0312	0,1412	0,0250	0,6893	0,6972	0,6723***
Self rated health, really good=0						
Good	0,7985**	0,5498*	0,1446**	0,4972	0,7131	0,5685
Passing	2,8310*	1,1974*	0,3509*	3,1824*	4,1539*	3,8030*
bad & really bad	10,0310*	3,6498*	1,4018*	28,8864**	43,4457*	42,0439*
> 6 months illness/consequences of accidents etc.	-4,7242*	0,0117	-0,0062	-3,7619**	-11,3847**	-11,3784**
Age	-0,0069	-0,0216	-0,0062**	-0,0225	-0,0192	-0,0131
Male (female=0)	-0,0252	-0,4939	-0,0802	-0,5946	-1,2437	-1,1635
Children living at home						
1 child >13 years (no children >13=0)	-0,5481	0,3375***	0,2934***	0,6571	0,5270	0,2337
2 children > 13 years	0,0874	0,5905*	0,2454*	0,7892	0,3108	0,0654
> 2 children > 13 years of age	-0,2570	0,6342**	0,0574	0,9330	-0,5766	-0,634
Education						
Skilled worker (unskilled =0)	-0,9199	-0,4426	-0,0535	-1,1360	-3,1934	-3,1399**
Semi-skilled	1,2052	0,0004	0,0321	2,9042	0,9715	0,9394
Junior college	-1,0596	-0,1160	0,0390	-0,6387	-1,0804	-1,1194
College	-0,8087	-0,1604	0,0737	-0,4355	-1,7419	-1,8156
University	-0,4301	0,2140	0,1028	-0,1204	-1,9876	-2,0903
Mics.	-3,474	-0,8897	-0,0110	-1,4073	-2,8634	-2,8524
Gross income (< 100.000 =0)						
100-199,000	2,6552	0,3837	0,1892	1,2891***	1,0519	0,8628
200-299,000	3,3201**	1,1261*	0,3169**	4,2280*	2,3532*	2,0363
300-399,000	3,8114*	1,4187*	0,3348*	2,8246*	3,9404***	3,6056***
400-499,000	4,1615*	0,7855*	0,1890	3,0901*	3,6700	3,4810
500-599,000	3,9645*	1,1425*	0,2809	3,1705**	3,0866	2,8057
600-699,000	2,9171**	0,7086*	0,1465	2,0620	2,6121	2,4656
700-799,000	4,3168*	0,3992	0,1726	1,8029*	2,6160	2,4435
800,000 and upward	3,4971	0,6652*	0,1329	2,0244**	2,7246	2,5917
do not wish to reveal/don't know	3,2549**	0,7571*	0,1899	2,4131*	2,9816***	2,7917

Form of remuneration	**Ord OLS**	**Median (quantile)**	**q25 (quantile)**	**q90**	**q95**	**q25-q95 quantile difference-**
By the hour (salaried =0)						
Base pay and performance related/piece work	1,0504	-0,2710	-0,0781	-0,1675	0,4250	0,5031
No management responsibility (yes=0)	-1,0647***	-0,2612***	-0,1009	-0,9924	-0,5311	-0,4302
Characteristic of place of work	1,4221**	0,5277*	0,0856	0,6786**	0,9922	0,9066
Public-private:						
Public sector (private sector=0)						
Public company	-0,6606	0,1704	0,1172	-0,2737	-0,0035	-0,1207
Other/don't know	0,0105	-0,2820	0,0117	0,8334	2,2089	2,1972
No. of employees	-0,3575	0,3782	-0,1810	0,7402	0,1273	0,3083
5-9 employess (1-4=0)						
10-19	2,3576***	0,3741	0,0716	1,1193	1,8514	1,7798
20 - 49	0,4099	0,2504	0,0047	-0,1734	3,0161	3,0114
50 - 99	0,8054	0,1037	-0,1169	-0,0314	1,7322	1,8491
100 - 249	0,8918	0,6434	0,0831	0,4716	3,1170	3,0339***
250 - 499	3,2323	0,5693	-0,0102	1,7135	3,6959	3,7061
> 500	1,4697**	0,5984**	-0,0096	1,1163	3,5236	3,5332**
don't know	0,6887	0,3066	-0,0169	0,4780	2,5110	2,5279***
Healht promotion at work	-1,2991	0,5316	-0,1679	-1,5090	-1,7002	-1,5323
No health promtion (yes=0)						
don't know	-0,2331	-0,0666	-0,0248	0,5065	0,7764	0,8013
Written policy on sicness absence	-0,0793	-0,3761	0,0326	-0,6334	0,8642	0,8316
No (yes=0)						
Don't know	-0,3621	-0,2075	-0,1076	0,1570	1,6170	1,7247***
Ever called to interview about sickness absence (yes=0)	-0,5797	-0,0643	-0,0073	0,7366	2,4133	2,4206**
No						
Overall satisfaction with place of work	-7,6645*	-3,4275*	-2,1795*	-12,0826*	-29,6653**	-27,4858*
Rather satisfied (very =0)						
Satisfied	-0,6583	-0,0028	-0,0074	0,3339	-0,0507	-0,04323
Dissatisfied	-0,0385	0,1860	0,0728	0,2957	-0,2341	-0,30689
Rather and very dissatisfied	2,3756***	0,6716**	0,1927	5,0589**	4,9273*	4,7346**
Physicially demanding type of work (always=0)	5,0109**	0,7678***	0,0629	4,9788	17,5853	17,5224
often						
now and then	-1,4704	-0,1798	0,0708	-0,846611	-1,4631	-1,5339
never	-2,4739	-0,1533	0,1034	-1,3295	-3,1476	-3,2510
	-1,6301	0,0436	0,0914	-1,0039	-2,2185	-2,3100

Work piles up if absent (always=0=	Ord OLS	Median (quantile)	q25 (quantile)	q90	q95	q25-q95 quantile difference-
_Often						
now and then	0,3017	0,4332	0,1279	-0,8702	-1,1953	-1,3233
never	-2,0976***	-0,1288	0,0156	-2,7999	-4,0882**	-4,1038
_Often	-2,4450*	-0,3483	-0,0478	-3,1832	-4,2180**	-4,1702
now and then	-3,1626	-0,9190	-0,2689	-8,0941	-9,6055	-9,3366
never	-1,6210	-0,9620	-0,2786	-8,5673	-9,7123	-9,4337
No. of physician contacts past 12 month	-2,2922	-0,9383	-0,2576	-8,8555	-10,4075	-10,1499
No of A&E and outpatient contacts past 12 month	1,8358*	0,6272*	0,1930*	1,5774*	1,8945*	1,7015*
Hospitalized (no=0)	1,2816***	1,1910*	0,3975*	3,8825*	4,2044*	3,8069*
Healt insurance	14,1690*	3,8995*	1,9799*	34,1318*	62,1702*	60,1903*
No (yes=0))						
Don't know	-0,1172	-0,2251***	-0,0509	-0,1236	0,3523	0,4031
Made use of health insurance past 12 months (pulled from a separate analysis)	1,0534	0,0015	-0,0652	0,7515	2,7111	2,7763
No (yes =0)						
Do not hold health insurance	*-2,6412*	*0,3956*	*-0,0177*	*-2,2436*	*-2,3169*	*-2,1396*
_cons	*-1,9845*	*0,467*	*-0,2003*	*-2,1579*	*-1,1030*	*-0.9027*
R2	13,0024	4,5518	2,1420	28,0412	56,7680	54,6260
	0.17	0.09	0.03	0.25	0.35	

* p<= 1%; ** p<=5%, *** p<=10%
documented in stata/kvartil-tabel and interkvartil (STATA log-files).

There are two striking results in table 7: The significant results for the self-assessed health status variable, showing, not surprisingly, that persons with bad or very bad health status had significantly more days of absence compared to person with very good health. The use of health services – private and public – is measured by three variables: Physician contact, outpatient contacts, and hospitalizations. All three are significant across the regressions. The sign – positive – may surprise some, but really should not. The use of services naturally co-varies positively the increased length of sickness absence.

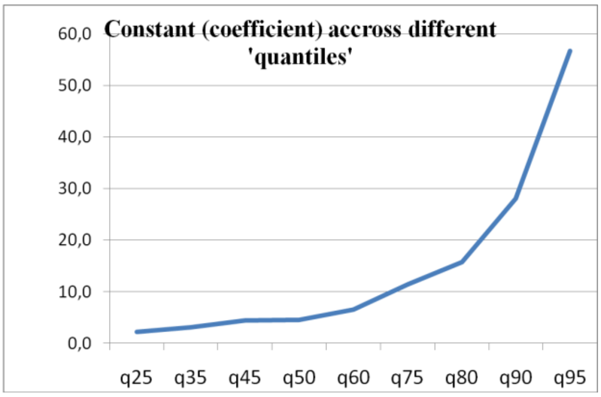
In table 7 the personal income is also interesting, but the reason is in part an unfortunate choice of reference group (the lowest income group with many part time workers). This will be changed in a revised version.

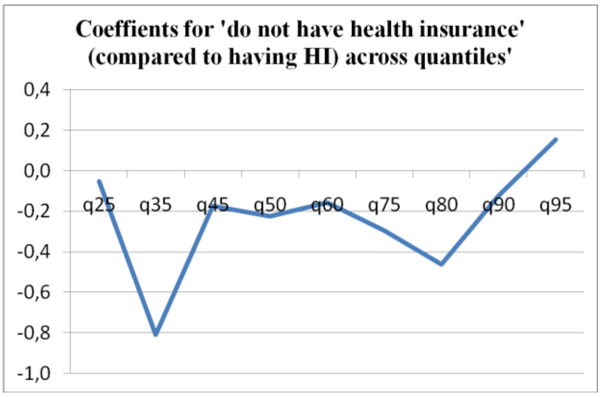
Table B in appendix I shows the results based on the HIS-dataset. This data set does not have as rich company details as the PRS material. Nevertheless, the results largely conform to the results in table 7, in particular as regards the health insurance variables.

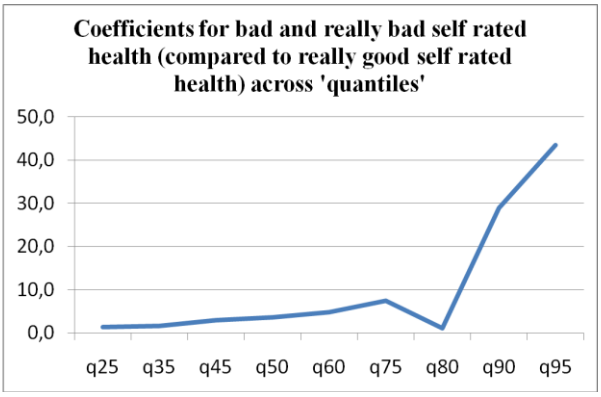
A good graphic representation is to look at the size of selected coefficient across a range of quantiles. Not only does it give a feel of what is achieved by using quantile regression, but also how things change across the (conditional) quantiles. Four examples are presented in figure 5: the constant, the 'not having HI', the 'not having management responsibility' and 'having bad or really bad self rated health'.

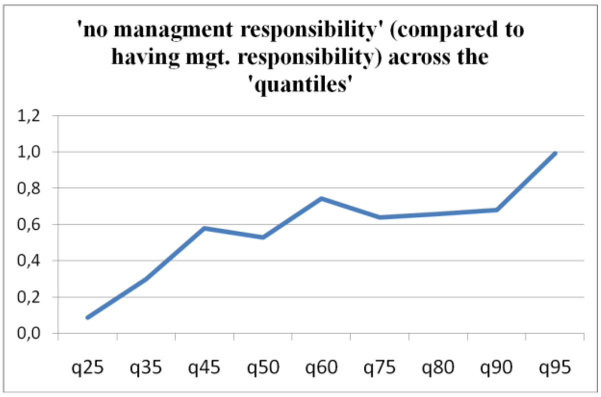
The constant clearly mirrors the move from few days of absence at the low end of the (conditional) distribution to considerably more days as we move towards the high end – indicating how the distribution is systematically investigated by using quantile regression (in looking at size of the coefficients recall that the omitted categories are (also) captured by the constant).

Figure 5: Graphs of the numerical size of coefficients across conditional quantiles for selected variables, PRS data (documented in kvartil-koefficienter) **.**









(the y-axis is measured in terms of days of sickness absence)

Turning to the 'do not have HI' – and recalling that the coefficients reported in table 7 were not significant – it is noteworthy, nevertheless, that towards the high end of the distribution there is a tendency for those not having HI to have more days of absence than those with HI.

In order to illustrate the behavior of two other coefficient, the health status variable and the (no) management responsibility, have also been graphed. The hypothesis behind including the management variable is that those with management responsibility have a higher ‘threshold’ before calling in ill. Looking at the health status variable it is clearly seen that those with bad and really bad health status compared to those with really good health status had considerably more days of absence that the high end of the distribution – stressing the obvious that health status matters. As for the management variable two observations are relevant: Those without management responsibility systematically have more days of absence compared to those with – and the tendency is increasing towards the high end of the distribution.

Finite mixture model (latent class model)[xxx]

A finite mixture model is a possible alternative or, more likely, complement to the quantile regression model in that it also explores different parts of a distribution, namely based on a latent groups, but based on a different philosophy than the quantile regression model.

In the health econometrics of the demand for health care it is common to use the two-part model (the hurdle model), TPM. The first part of the TPM is a binary outcome model, typically a logistic model, that describes the distinction between non-users and users. The second part describes the distribution of use conditional on some use, modeled either as a continuous or integer-valued random variable. In the latter case typically the negative binomial model. The appeal of the TPM in part stems from the high incidence of zero usage (read: many without sickness absence).

The sharp dichotomy between users and non-users in the TPM has been challenged by Deb and Trivedi[71-72] and, among others, Bago[73]. They claim that a more tenable distinction for typical cross-sectional data may be between an “infrequent user” and a “frequent user” of medical care – or in the present case short and long term sickness absence - the difference being determined by health status, attitudes to health risk, and choice of life-style. In their proposed alternative model, the latent class model, LCM, there is no distinction between users and non-users of care. Instead there is a distinction between groups with high average demand and low average demand based on two latent classes.

Deb and Trivedi hypothesized that the underlying unobserved heterogeneity which splits the population into latent classes is based on an individual’s latent long-term health status. Proxy variables such as self-perceived health status and chronic health conditions may not fully capture population heterogeneity from this source. Consequently, in the case of two latent subpopulations, a distinction may be made between the “healthy” and the “ill” groups, whose demands for medical care (or sickness absence) are characterized by low mean and high mean, respectively.

The mixture/latent class approach can be interpreted as allowing for latent groups/classes in the population. The data for each group may be characterized by a parameter vector – the same for both groups. Since the group to which an individual belongs is not observed directly, a mixing

[xxx] I am grateful to Nabanita Datta Gupta for suggesting this approach which in no way is pursued in detail here. In essence it is only scratching the surface in an exploratory manner.

probability is used to classify individuals probabilistically. The mixture negative binomial model has the virtue of being conceptually simple In the following estimation two latent groups are assumed (‘short’ and ‘long’ term sickness), cf. figure 4 which gives a degree of credence to this assumption, and the mixture model used is the negative binomial[XXXI].

Table 7A shows the results for the two specifications of health insurance situation, omitting all the control variables (see appendix): Use of insurance within the past 12 months and a dichotomous health insurance status variable like in the quantile regression analysis.

As regards use of health insurance compared to non-use and not having an health insurance at all the results are clear for both components (low mean sickness absence (component2) and high(er) mean absence, component 1. This is judged from the constant term, cf.appendix I, table BB and BBB): There is no significant difference with respect to sickness absence for the group that used their health insurance compared to the non-user, but the non-insurance holders had a significantly lower sickness absence compared to the group that used their health insurance.

Table 7A: Results from two-class latent model with negative binomial as mixture. N=3997

	Coef.	**Std. Err.**	**P>|z|**	**[95% Conf. Interval]**		**Coef.**	**Std. Err.**	**P>|z|**	**[95% Conf. Interval]**	
Used health insurance past 12 months	**component1 (0.33)**					**component2 (0.67)**				
No (yes =0)	-0.4160	0.2060	-2.02	-0.8196	-0.0123	-0.0898	0.0734	-.221	-0.2336	0.0540
Do not hold health insurance	-0.1761	0.1954	-0.9	-0.5591	0.2069	-0.2345	0.0752	-.002	-0.3819	-0.0870
Healt insurance										
Yes (no=0)	-0.0802	0.1549	0.604	-0.3838	0.2233	0.1608	0.0537	0.003	0.0555	0.2660

The full regression results are found in appendix I, table II and III. The same independent variables as in table 7.

Using the dichotomy ‘holds – does not hold health insurance’ there is an interesting result in that the there is a significant difference between insurance and non-insurance holders for component 2 (low mean level of absence) in that those with health insurance actually have higher absence than those without insurance. For long term sickness absence there was no significant difference.

Overall then – and given that the mixture analysis primarily is explorative – the results confirm the results from the quantile regression analysis in that health insurance does not appear to be associated with lower sickness absence.

[XXXI] STATA’s version 10 fmm routine is used, negbin1 is used for mixture.

The detailed results in appendix ??? show several cases of different effects of the independent variables (absolute size, sign and significance) for the two latent groups, e.g. the effect of number of chronic illnesses, age

Propensity score and matching estimator approach to the effect of health insurance on sickness absence

In a *causal sense* the question in relation to the results from the quantile and mixture regression above is: Has it been shown that 'health insurance' or 'having made use of health insurance' *does not cause* a reduction in (long term) sickness absence – despite the non-significant results? This is not the place for a full discussion of the causal interpretation of regression analysis, see for instance Angrist and Pischke, section 3.2[63] or Morgan and Winship[64] chapter 5. A few observations are in place, however, as transition to the section on propensity score and matching estimators where causality under certain assumptions may be claimed, but note, however, that under 'certain' assumptions this also holds for regression analysis, e.g. a fully saturated model. So, the propensity-matching approach is not in all regards superior causal-wise to OLS:.

It should be noted that the two 'intervention variables' in table 7 and 7A, most likely are exogenous and endogenous respectively. 'Health insurance' in essence is not a choice variable for the individual employee, but a given – decided by the employer - and reasonably must be assumed independent of the outcome variable (sickness absence) whereas 'use of health insurance' is a choice variable for the individual employee, i.e. there is undoubtedly self selection. The question of the conditional independence assumption or put differently selection on observables has been attempted through the use of an extensive set of relevant covariates. There are 'good' and 'bad' control variables[63]. Angrist and Pischke note that 'bad' controls might as well be dependent variables too, while 'good' controls are variables that we can think of as having been fixed at the time the regressor of interest was determined. A look at the set of control variables in table 7 and table B points towards largely 'good' controls. If, on the other hand, actual use of health services had been used as a control, it would have been an example of a 'bad' control. However, whether there is omitted variables[XXXII] – and hence omitted variable bias – cannot be decided definitively[XXXIII]. Overall, it probably would not be unreasonable to use a causal interpretation of the analysis with 'health insurance' as the intervention variable.

[XXXII] We have abstained from using one or all of the standard tests for omitted variables: The three classic likelihood-based approaches are: *LR test*:estimate model with and without constraint ( ) = 0. *Wald test*: estimate the model without the constraint, and the *(effcient) Score/Lagrange Multiplier test:* estimate the restricted model.
[XXXIII] Angrist and Pischke say that "although simple, the OVB formula [omitted variable bias] is one of the most important things to know about regression. The importance of the OVB formula stems from the fact that if you can claim absence of omitted variable bias, then typically you're also saying that the regression you've got is the one you want. And the regression you want usually has a causal interpretation. In other words, you are prepared to lean on CIA [conditional independence] for a causal interpretation of the long regression estimates"[63], p. 62.

propensity score

An increasingly popular approach to estimate the effect of (a binary) 'treatment' variable, e.g. 'having-not having health insurance' or 'used – not used health insurance to obtain health services'. is to estimate a propensity score (propensity/probability of being 'assigned' to a particular (binary) situation given a set of observable co-variables) of receiving 'treatment') and use the propensity score to match 'treated' and 'untreated' using a variety of matching techniques (nearest neighbor, kernel etc) .

Matching is based on the assumption that the outcomes, y_0 y_1 are independent of treatment, D, given a set of observed covariates, *x*:

(1) $\boldsymbol{y_0, y_1 \perp D | x}$

Rosenbaum and Rubin[65] showed that given the propensity score, the conditional probability of receiving treatment given *x* is :

(2) $\boldsymbol{p(x) = P[D = 1 | X = x]}$

and assuming covariate balance:

(3) $D \perp x | p(x)$

then (1) implies:

(0) $\boldsymbol{y_0, y_1 \perp D | p(x)}$

Meaning, therefore, that it is sufficient to match on the propensity score

There are two issues concerning propensity scores: choice of co-variates and statistical issues[66].

Ideally the choice of co-variates in the propensity score-equation should be theory-driven, i.e. in the present case based on the demand for health insurance. It is all but too rare, however, to see references to this obvious point. Therefore, in practice it boils down to – not only data availability, but common sense – and paying too little attention to (relevant/likely) omitted unobservable variables that may create bias[XXXIV]. On the one hand it is obvious, on the other hand an important limitation, in particular because in some quarters it is almost believed that propensity score matching is (almost) as good as truly randomized experiments with clear assignment of treatments. Rubin[67] notes the obvious, but all too often overlooked: "... It is important to keep in mind that even propensity score methods can only adjust for observed confounding covariate and not for the unobserved one. This is always a limitation of non-randomized studies compared with randomized studies, where the randomization tends to balance the distribution of all covariates, observed and unobserved", p. 762.

As to statistical issues in estimating the propensity score, Guo and Fraser ask[66]: "What defines the 'best' logistic regression? The answer is simple: We need propensity scores that balance the two

[xxxiv] In particular when – as here – one is using person data, where the choice of providing health insurance, i.e. 'treatment', in essence is made by the employer, not the individual employees. There is most likely to be important unobservable variables at the employer-level that most people related observable covariates do not capture. In this regard the PRS data are far better than the HIS data, i.e. a number of variables capture the work environment – and hence type of company.

groups on the observed covariates. By this criterion, a best logistic regression is a model that leads to estimated propensity scores that best represent the true propensity scores. The challenge is that the true propensity scores are unknown, and that therefore, we must seek methods to measure the fit between the estimated and the unknown true scores" (p. 138). They go on to stress – as noted above - the importance of including models carefully chosen and appropriate conditioning variables in the correct. They also note that estimates of treatment effects are sensitive to different specifi-cations of conditioning variables.

A 'good' logistic regression model should minimize the overall sample prediction error. This may be done by using the boost-routine in Stata. This has, however, not been done in the current analysis.

The intervention can be either binary treatment conditions, e.g. 'has – does not have health insurance' with the propensity score derived from the predictions from a logistic regression. However, in many real world situations there are more than two treatment conditions (treatment dosage). Either a continuum of treatment, e.g. health care services, or polychotomous treatment conditions. This is relevant in the present context in that – as discussed earlier – the real issue is not having health insurance but rather a) has it been used (in connection with sickness absence) and b) is positive: how much, cf. tables 2,3, and 4 above. Table 8 shows the treatment categories

Treatment dosage can be modeled in two ways. One may estimate a single scalar propensity score using ordered logistic regression, and then match on the scalar propensity score and proceed as for two treatment groups. The alternative is to estimate a propensity score for each level of the treatment dosage using a multinomial logistic regression, and then define the inverse of a particular estimated propensity score as sampling weight to conduct a multivariate analysis of outcome. Guo and Fraser essentially recommends what they term 'subset analysis', i.e. subsets of the data corresponding to the does categories, cf. table 8. '…in our opinion, conducting an efficacy subset analysis is an efficient and viable alternative to modeling doses of treatment', p. 167[66]

Table 8: Treatment categories in the PRS data

	N	%	Cum.
Do not hold HI	2,488	61.28	61.28
Hold HI, but not used it	1,158	28.52	89.80
Used, but not for absence	281	6.92	96.72
Used only for absence	40	0.99	97.71
Used partly for absence	80	1.97	99.68
Do not remember	13	0.32	100.00

Disregarding the last category, it seems natural to use ordered logistic regression in this case because the 'ordering' does reflect a rank order.

Propensity scores and treatment effects for the binary intervention, 'has – does not have HI', have been estimated for both data sets[XXXV]. The results for the HIS have been relegated to Appendix I, Tables C to F.

The results for the PRS analyses are shown in tables 9-12, with table 10 showing the results of particular interest, namely treatment effects of HI – and showing no statistically significant effect of health insurance.

Table 9: Propensity score for health insurance (1=success), PRS data set

Characteristics of the employees	Odds Ratio	Robust Std. Err,	z	P>|z|	95% Confidence, Interval	
Health						
Number of chronic diseases	0,9970	0,0413	-0,070	0,941	0,9192	1,0813
Self rated health, really good=0						
Good	0,9708	0,1159	-0,250	0,804	0,7682	1,2268
Passing	0,7293	0,1066	-2,016	0,031	0,5476	0,9713
bad & really bad	0,6924	0,1770	-1,044	0,150	0,4195	1,1428
> 6 months illness/consequences of accidents etc0,	0,9133	0,1105	-0,750	0,453	0,7205	1,1577
Age	0,9859	0,0042	-3,037	0,001	0,9777	0,9941
Male (female=0)	0,9539	0,0871	-0,520	0,605	0,7976	1,1408
Education						
Skilled worker (unskilled =0)	1,1800	0,1974	0,990	0,323	0,8501	1,6380
Semi-skilled	1,0435	0,2213	0,200	0,841	0,6886	1,5812
Junior college	1,0319	0,1843	0,180	0,860	0,7272	1,4643
College	0,7929	0,1240	-1,480	0,138	0,5836	1,0772
University	0,9484	0,1595	-0,320	0,753	0,6821	1,3187
Mics0,	0,9503	0,2417	-0,200	0,841	0,5772	1,5645
Gross income (< 100,000 =0)						
100-199,000	1,4551	0,5816	0,940	0,348	0,6647	3,1850
200-299,000	2,7278	1,0281	2,066	0,008	1,3032	5,7100
300-399,000	4,6438	1,7317	4,012	0,000	2,2359	9,6447
400-499,000	5,3957	2,0529	4,043	0,000	2,5597	11,3737
500-599,000	5,4392	2,1743	4,024	0,000	2,4847	11,9070
600-699,000	7,3485	3,0904	4,074	0,000	3,2228	16,7558
700-799,000	5,2401	2,4159	3,059	0,000	2,1228	12,9351
800,000 and upward	8,0865	3,4137	4,095	0,000	3,5353	18,4968
do not wish to reveal/don't know	4,6665	1,7678	4,007	0,000	2,2209	9,8051

[XXXV] In STATA logistics, psmatch2, pstest, and rbounds have been used.

Type of remuneration	Odds Ratio	Robust Std. Err,	z	P>|z|	95% Confidence, Interval	
By the hour (salaried =0)	0,3686	0,0548	-6,071	0,000	0,2755	0,4933
Base pay and performance related/piece work	0,7089	0,1102	-2,021	0,027	0,5227	0,9615
Characteristic of place of work						
Public-private:						
Public sector (private sector=0)	0,0460	0,0056	-25,009	0,000	0,0362	0,0586
Public company	0,2561	0,0638	-50,470	0,000	0,1571	0,4173
Other/don't know	0,2367	0,1006	-30,390	0,001	0,1029	0,5446
No, of employees						
5-9 employess (1-4=0)	1,6223	0,3320	20,360	0,018	1,0862	2,4230
10-19	1,8220	0,3614	30,020	0,002	1,2351	2,6877
20 - 49	2,4977	0,4492	50,090	0,000	1,7557	3,5534
50 - 99	2,4404	0,4764	40,570	0,000	1,6646	3,5778
100 - 249	2,7520	0,5132	50,430	0,000	1,9095	3,9663
250 - 499	3,6321	0,8213	50,700	0,000	2,3318	5,6576
> 500	2,6497	0,4679	50,520	0,000	1,8745	3,7456
don't know	1,0847	0,5813	0,150	0,879	0,3795	3,1006
Healht promotion at work						
Sundhedsordning (yes=0)	0,1995	0,0201	-160,020	0,000	0,1638	0,2429
don't know	0,2705	0,0526	-60,720	0,000	0,1847	0,3960
Physicially demanding type of work (always=0)						
often	0,7342	0,1583	-10,430	0,152	0,4812	1,1202
now and then	0,7998	0,1585	-10,130	0,260	0,5424	1,1795
never	1,0154	0,2012	0,080	0,939	0,6885	1,4974
Work pile up if absent (always=0)						
_Often	1,5948	0,2938	20,530	0,011	1,1115	2,2884
now and then	1,3351	0,2164	10,780	0,075	0,9717	1,8344
never	1,1669	0,1896	0,950	0,342	0,8486	1,6045
Physically exhausting (always=0)						
_Often	1,6607	0,4856	10,730	0,083	0,9363	2,9457
now and then	1,5162	0,4345	10,450	0,146	0,8646	2,6588
never	1,5028	0,4393	10,390	0,163	0,8474	2,6654

(documented in C: stats/propensity-HI

Table 10 does not provide support for at treatment effect of health insurance on sickness absence.

Table 10: Treatment effects using propensity equation reported in table 8, PRS data

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
q46a (sickeness absence days)	Unmatched	5,81	6,55	-0,732	.613593392	-1.19
	ATT	5,80	6,14	-0,341	120.585.054	-0.28
	ATU	6,49	8,32	1,832	.	.
	ATE			1,006	.	.

Propensity balancing test, PSM , for each covariate are reported in detail in appendix I, table G. Table G shows that for but one variable (at the 5% level) there were no significant differences in the matched means of the variables.

Table 11 shows the summary of the balancing tests confirming the impression from table G

Table 11: Summary of the balancing tests

Sample	Pseudo R2	LR chi2	p>chi2
Unmatched	0.347	1839.51	0.000
Matched	0.010	36.12	0.852

Matching is based on the conditional independence assumption, meaning that the researcher should observe all variables simultaneously influencing the participation decision and outcome variables. This is a strong identifying assumption. Hence, checking the sensitivity of the estimated results with respect to deviations from this identifying assumption is of great importance.

If there are unobserved variables which simultaneously affect assignment into treatment and the outcome variable, a ‘hidden bias’ might arise to which matching estimators are not robust. Since it is not possible to estimate the magnitude of selection bias with non-experimental data, this problem can be addressed with the bounding approach proposed by Rosenbaum and implemented as STATA routine by Becker and Caliendo (mhbounds)[68] for a binary outcome variable and by Gangl for a continuous outcome variable[69] (rbounds – which is used below).

The basic question asked is whether or not inference about treatment effects may be altered by unobserved factors. In other words, one wants to determine how strongly an unmeasured variable must influence the selection process in order to undermine the implications of the matching analysis. This of course, is not a test of the unconfoundedness assumption itself, because this would amount to testing that there are no (unobserved) variables that influence the selection into treatment. Instead, Rosenbaum bounds provide evidence on the degree to which any significance results hinge on this untestable assumption. Clearly, as noted by Rosenbaum, Becker and Caliendo and Gangl, if the results turn out to be very sensitive, the researcher might have to think about the validity of his/her identifying assumption and consider alternative estimation strategies.

Guo and Fraser (chapter 8)[66] nicely summarizes the issue of selection bias in relation to propensity score and matching estimators. At the general level a valid application of these approaches requires broad knowledge and skill in regard to at least five points. 1. a thorough understanding of the sources of selection bias, 2.conducting a careful investigation of existing data and literature to identify all possible covariates that might affect selection and be used as covariates in for instance generating the propensity score (but similarly in the regression approaches). 3. develop an understanding of the fit between the data generation process and the assumptions in the propensity and (in particular) matching estimator model. 4. provide cautious interpretations the findings in view of the assumptions, finally 5. conduct sensitivity analysis to gauge the level of sensitivity of findings to hidden bias.

Table 12 shows the *preliminary* results of the Rosenbaum sensitivity analysis as implemented in rbound in Stata.

Table 12: (preliminary) Sensitivity analysis (Rosenbaum), PRS data

Rosenbaum bounds for **delta** (N = **1365** matched pairs)

Gamma	sig+	sig-
1	.026934	.026934
1.05	.122093	.003547
1.1	.330896	.00031
1.15	.60147	.000019
1.2	.821812	8.2e-07
1.25	.940699	2.7e-08
1.3	.98517	6.7e-10
1.35	.997164	1.3e-11
1.4	.999577	2.1e-13

* gamma - log odds of differential assignment due to unobserved factors
sig+ - upper bound significance level
sig- - lower bound significance level

Table 12 reports p-values from Wilcoxon signed ranks test for the averaged treatment effect for the treated while setting the level of hidden bias to a certain value of gamma. sig+ and sig+ refers to maximum and minimum of the p-value. Gamma embeds the assumption about endogeneity in treatment assignment in terms of the odds ratio of differential treatment assignment due to an observed covariate[XXXVI]. By comparing the Rosenbaum bounds on treatment effects at different

[XXXVI] Following the interpretation by DiPrete and Gang[69]. A similar interpretation is found in section 8.4 and 8.5 of Guo and Fraser[66] levels of gamma it is possible to assess the strength such unmeasured influence must have in order that the estimated treatment effects from propensity score matching would have arisen purely through nonrandom assignment.

From the second column it is seen that already for gamma at 1.05 we have to question the conclusion in table 10, i.e. it is attained if an unobserved covariate caused the odds ratio of treatment assignment to differ between treatment and control cases by a factor of about 1.05.

Summary, conclusions and discussion

From a Danish policy perspective there is considerable interest tied to trying to determine whether or not holders of employer paid health insurance experience lower sickness absence, in particular long term sickness. This is no trivial task, however. The following points underscore this – and also illustrate the strong and weak points of the present analysis.

However, before delving into discussion points it should be noted that it has not been possible to show a sickness reducing effect on sickness absence regardless of econometric approach: Quantile regression, finite mixture model, or matching.

First of all, theoretical models not only of sickness absence but also of how health insurance is likely to influence sickness absence are scarce, almost non-existent, in part because the issue is only particularly relevant in an institutional setting like the Danish[XXXVII]. Here, two modeling approaches have been used: one for sickness absence and (less rigorously) a moral hazard and access to health care interpretation of health insurance.

In models of health insurance the issue is ex post moral hazard, i.e. 'overconsumption' of health care due to HI and/or an access reason for holding HI. There is only scant evidence supporting the existence of ex post moral hazard for Danish employer paid HI, and hence a higher level of consumption of health care that ideally should subsequently generate a lower level of long term sickness absence[XXXVIII]. The access issue is about a) faster access to treatment or medical certification, e.g. less waiting time, and b) possibly better coordinated care. As regards point a) 60% of the respondents in the PSR data indicate that this is the case in their experience, table 6A.

[XXXVII] But see for instance BUPA's homepage where (reduction) of sickness absence plays a prominent role: http://www.bupa.co.uk/business/all-business/business-health-insurance-2/health-insurance-benefits. BUPA is the biggest British health insurance company. BUPA also operates internationally – and has diversified into, among other things health care facilities. In Denmark BUPA owns IHI.

[XXXVIII] There is a logical issue here, in particular as regards the use of cross sectional data: Sickness absence (in some cases) leads to use of health services , i.e sickness absence causes use of health services. The question then is whether this utilization in a later period reduces sickness absence, for instance because of an operation in period 1. This dynamic and relevant issue cannot be addressed with cross section data unless utilization and reasons for use are tracked to sickness absence using retrospective questions. Data of this nature has not been available for the analyses presented here.

The discussion of ex post moral hazard clarifies that it cannot be HI per se that may generate lower sickness absence, but the actual use of services, i.e. HI at best is ‘enabler’ and hence only has an indirect link to sickness absence that require diagnosis or treatment.

Another issue is modeling sickness absence (regardless of HI). To this end a ‘demand’ function for sickness absence was developed based on explicit inclusion of health status and work environment. Eq. (5) was part of the reasoning behind the regression specification of determinants of sickness absence.

The second point concerns specification of empirical models – a specification that, based on observable variables, should minimize selection bias. This requires two things: theory and knowledge of the field and then relevant data. As regards the first the present analysis not only presents theoretical reasoning but also uses epidemiological results.

The third issue then is relevant data. Two survey data sets have been used here. One of them, the PSR data, has several detailed questions addressing sickness absence and health insurance or use of health insurance. The PSR data set illustrates this. For instance, if the answer to use of HI within the past 12 months, the next question was: is it related to sickness absence. Without this supplementary question one might mistakenly conclude all use of HI as related to sickness absence.

The fourth issue concerns choice of econometric models. A core issue here is the possibility of interpreting the intervention variable in a causal sense. In other words, is it permissible to claim that HI (or use of HI) ‘causes’ lower sickness absence among HI-holder compared to non-holders. This in turn leads to selection based on observable covariates. Here the issue becomes how well do ‘corrective methods’, i.e. propensity score or ordinary OLS handle selection bias and what are the embedded assumptions. For the camp that swears to propensity score + matching estimator, the question is the sensitivity to (possible) hidden bias from unobserved variables that are associated both with assignment to treatment, i.e. HI, and the outcome variable, i.e. sickness absence. Propensity-score matching estimators are not consistent estimators for treatment effects if the assignment to treatment is endogenous – in other word variables that affect the assignment process are also related to the relevant outcome. Specifically, do firms who offer HI to employees (assignment) also have lower/higher sickness absence? This points the possibility of an important source of hidden bias due to unobserved variables.

No definitive judgment has been passed on this discussion. Instead both approaches have been used and a first go at testing for bias due to unobserved variables has been put in place. For the PSR data they lead to the same conclusion, namely absence of effect of HI on sickness absence. As regards the HSR data the results are conflicting.

The host of issues related to the estimation of the effect of a) HI and b) use of HI on sickness absence as is often the case with research raises more questions and answers. However, it is important to note that the issue can never be reduced to an econometric one: Common sense (is it really HI per se that …) and theory are important components in selection of covariates.

Bibliography

1. Folketinget, Parliament. Legislative proposal L 97, comments on proposal, second session 2001/2002. Proposal in Parliament (Folketinget) 2001.
2. Colombo F, Tapay N. Private Health Insurance in OECD Countries: The Benefits and Costs for Individuals and Health Systems. Towards high-performing health systems: policy studies. Paris: OECD; 2004.
3. OECD. Proposal for a taxonomy of health insurance (OECD Study on Private Health Insurance). Paris: OECD; 2004.
4. Borchsenius V, Hansen JV. Er sundhedsforsikrede mindre syge end uforsikrede? Forsikring & Pension Analyserapport 2010:6 2010.
5. Pedersen KM, Arendt JN, Kiil A. Arbejdspladserne og sundhedsforsikring. Foreløbige resultater fra spørgeskemaundersøgelse af syge-og sundhedsforsikringer. Draft, upcoming Health Economics working paper - essentially completed September 2009 2010.
6. Statistics Denmark. Statistikbanken/Stat Bank Denmark (self generated tables fra 'labour market/statistics on absence/FRA01 - and similar tables FRA02-FRA08). 2011.

Ref Type: Online Source

7. Johansen K, Andersen JS, Mikkelsen S, Pass O, Raffnsøe S, Lynge E. Controlling sickness absence: A study of changes in the Danish sickness absence legislation since 1973. Health Policy 2010;86(1):109-118.
8. Lusinyan L, Bonato L. Work Absence in Europe. IMF Staff Papers 2007;54(3):475-538.
9. Beskæftigelsesministeriet. Analyse af det danske sygefravær. København: Beskæftigelsesministeriet; 2003.
10. Beskæftigelsesministeriet. Analyse af sygefraværet. København: Beskæftigelsesministeriet; 2008.
11. Pedersen KM. Langvarigt sygefravær: udfordringer og potentiale. In: Beskæftigelsesrådet, editor. Beskæftigelsesrådets beskæftigelsespolitiske redegørelse . København: 2008:15-31.
12. Nord-Larsen M. Det langvarige sygefravær. København: SFI; 1989.
13. Nord-Larsen M. Langvarig sygdom : problemer og konsekvenser. København: Socialforskningsinstituttet, SFI; 1991.
14. Kilsgaard J, Sherson D, Bültman U, Olsen J, Pedersen J, Lund T. Koordineret Indsats for Arbejdsfastholdelse (KIA) af sygemeldte med længerevarende smertegener fra bevægeapparatet. København: Arbejdsmiljøinstituttet; 2006.
15. KAD, SID. Hvorfor vente? Initiativer for hurtiger og bedre indsats for sygemeldte. 2002.
16. Swedish Councilon Technology Assessment in Health Care S. Sickness absence: causes, consequences, and physicians' sickness certification practice. A systematic literature review by the Swedish Council on Technology Assessment in Health Care). Scand J Public Health Suppl 2004;63:3-263.
17. OECD. Sickness, Disability and Work - breaking the barriers. A synthesis of findings across OECD countries. Paris: OECD; 2010.
18. Gilleskie DB. A dynamic stochastic model of medical care use and work absence. Econometrica 1998;66(1):45.
19. Granlund D. The effect of health care expenditure on sickness absence. Eur J Health Econ 2010;11:555-568.
20. Granlund D. "Waiting for the other shoe to drop": Waiting for healtyh care and duration of sick leave. Working Paper 8/2010 Örebro University 2010.
21. Kjellberg J, Andreasen MN, Søgaard J. Private sundhedsforsikringer - notat udarbejdet for LO. København: DSI - Dansk Sundheds Institut; 2010.
22. Willert MV, Thulstrup AM, Bonde JP. Effects of a stress management intervention on absenteeism and return to work - results from a randomized wait-list controlled trial. Scand J Work Environ Health 2010.

23. Ektor-Andersen J, Ingvarsson E, Kullendorff M, Orbaek P. High cost-benefit of early team-based biomedical and cognitive-behaviour intervention for long-term pain-related sickness absence. J Rehabil Med 2008;40(1):1-8.
24. Linton SJ, Boersma K, Jansson M, Svard L, Botvalde M. The effects of cognitive-behavioral and physical therapy preventive interventions on pain-related sick leave: a randomized controlled trial. Clin J Pain 2005;21(2):109-119.
25. Landstad B, Vinberg S, Ivergard T, Gelin G, Ekholm J. Change in pattern of absenteeism as a result of workplace intervention for personnel support. Ergonomics 2001;44(1):63-81.
26. Carroll C, Rick J, Pilgrim H, Cameron J, Hillage J. Workplace involvement improves return to work rates among employees with back pain on long-term sick leave: a systematic review of the effectiveness and cost-effectiveness of interventions. Disabil Rehabil 2010;32(8):607-621.
27. Bang CK, Lund T, Labriola M, Villadsen E, Bultmann U. The fraction of long-term sickness absence attributable to work environmental factors: prospective results from the Danish Work Environment Cohort Study. Occup Environ Med 2007;64(7):487-489.
28. Borritz M, Christensen KB, Bultmann U et al. Impact of burnout and psychosocial work characteristics on future long-term sickness absence. Prospective results of the Danish PUMA Study among human service workers. J Occup Environ Med 2010;52(10):964-970.
29. Bultmann U, Rugulies R, Lund T, Christensen KB, Labriola M, Burr H. Depressive symptoms and the risk of long-term sickness absence: a prospective study among 4747 employees in Denmark. Soc Psychiatry Psychiatr Epidemiol 2006;41(11):875-880.
30. Christensen KB, Labriola M, Lund T, Kivimaki M. Explaining the social gradient in long-term sickness absence: a prospective study of Danish employees. J Epidemiol Community Health 2008;62(2):181-183.
31. Eshoj P, Jepsen JR, Nielsen CV. Long-term sickness absence - risk indicators among occupationally active residents of a Danish county. Occup Med (Lond) 2001;51(5):347-353.
32. Labriola M, Lund T, Christensen KB et al. Does self-efficacy predict return-to-work after sickness absence? A prospective study among 930 employees with sickness absence for three weeks or more. Work 2007;29(3):233-238.
33. Lund T, Labriola M, Christensen KB, Bultmann U, Villadsen E, Burr H. Psychosocial work environment exposures as risk factors for long-term sickness absence among Danish employees: results from DWECS/DREAM. J Occup Environ Med 2005;47(11):1141-1147.
34. Lund T, Labriola M, Christensen KB, Bultmann U, Villadsen E. Physical work environment risk factors for long term sickness absence: prospective findings among a cohort of 5357 employees in Denmark. BMJ 2006;332(7539):449-452.
35. Lund T, Labriola M, Villadsen E. Who is at risk for long-term sickness absence? A prospective cohort study of Danish employees. Work 2007;28(3):225-230.
36. Stoltenberg CD, Skov PG. Determinants of return to work after long-term sickness absence in six Danish municipalities. Scand J Public Health 2010;38(3):299-308.
37. Christensen KB, Lund T, Labriola M, Bultmann U, Villadsen E. The impact of health behaviour on long term sickness absence: results from DWECS/DREAM. Ind Health 2007;45(2):348-351.
38. Labriola M. Conceptual framework of sickness absence and return to work, focusing on both the individual and the contextual level. Work 2008;30(4):377-387.
39. Brown S, Sessions JG. The Economics of Absence: Theory and Evidence. Journal of Economic Surveys 1996;10:23-53.
40. Allen SG. An empirical model of worker attendance. Review of Economics and Statistics 1981;71(1):1-17.
41. Barmby T, Sessions JG, Treble JG. Absenteeism, Efficiency Wages and Shirking. Scandinavian Journal of Economics 1994;96(4):561-566.
42. Shapiro C, Stiglitz JE. Equilibrium Unemployment as a Worker Discipline Device: Reply. American Economic Review 1985;75(4):892-893.
43. Barmby T, Ercolani M, Treble J. Sickness Absence: An International Comparison. Economic Journal 2002;112:F315-F331.

44. Chatterji M, Tilley CJ. Sickness, Absenteeism, Presenteeism, and Sick Pay. Oxford Economic Papers 2002;54(4):669-687.
45. Ose SO. Working Conditions, Compensation and Absenteeism. Journal of Health Economics 2005;24(1):161-188.
46. Coles MG, Treble JG. The price of work reliablity. Economics Letters 1993;44:149-155.
47. AFSA C, GIVORD P. The Impact of Working Conditions on Sickness Absence : A Theoretical Model and an Empirical Application to Work Schedules. 20092009-06 - Serie des documents de travail du CREST. Institut national de la statistique et des etueds economicqes.)
48. Becker G. A theory of the allocation of time. The Economic Journal 2010;75(299):493-517.
49. Grosssman Michael. The demand for health: a theoretical and empirical investigation. NBER Occasional Papers 119, Columbia University Press, New York 1972 .

Ref Type: Generic

50. Christiansen T, Pedersen K, Harvald B, et al. A health economic investigation of the effect of treatment of hypertension. Odense: Institut for Sundhedsøkonomi og Sygdomsforebyggelse, Odense Universitet, 1987
51. Christiansen T, Pedersen KM, Harvald B, et al. An Investigation of the Effect of Regional Variation in the Treatment of Hypertension. 28. Soc. Sci. Med.[2], 131-139. 1988.

Ref Type: Generic

52. Currie J, Madrian B. Health, health insurance and the labor market. In: Ashenfelter O, Card D, editors. Handbook of Labor Economics, vol 3C. Amsterdam: 1999:3309-3416.
53. Gruber J. Health insurance and the labor maret. In: Culyer AJ, Newhouse JP, editors. Handbook of health economics, vol 1A. 2000:645-706.
54. Bundorf MK. Employee demand for health insurance and employer health plan choices. Journal of Health Economics 2002;21:61-88.
55. Glied S, Zivin G. Modeling Employer Decisions to Offer Health Insurance. Journal of Policy Analysis and Management 2004;23(2):281-313.
56. Nyman JA. The Value of Health Insurance: The Access Motive. Journal of Health Economics 1999;18(2):141-152.
57. Nyman JA. The theory of demand for health insurance. Stanford: Stanford University Press 2002;xiv, 201.
58. Kiil A, Arendt JN. Complementary private health insurance and the use of health care services: Identifying ex post moral hazard in the Danish market. Conference paper, Annual meeting of Nordic health economists, Umeå, August 2010 2010.
59. Kiil A. Does employment-based private health insurance increase the use of covered health care services? A matching estimator approach. Second year ph d-presentation, November 23, 2010 University of Southern Denmark 2010.
60. Severens JL, Mulder RJ, Laheij RJ, Verbeek AL. Precision and accuracy in measuring absence from work as a basis for calculating productivity costs in The Netherlands, *Social Sci Med* **51** (2000), 243-349. Social Sci Med 2000;51:243-249.
61. Koenker R. Quantile Regression. Cambridge and New York: Cambride University Press; 2005.
62. Koenker R, Hallock KF. Quantile Regression. Journal of Economic Perspectives 2001;15(4):143-156.
63. Angrist JD, Pischke J-S. Mostly Harmless Econometrics. An Empiricist's Companion. Princeton and Oxford: Princeton University Press; 2009.
64. Morgan SL, Winship C. Counterfactuals and causal inference. Methods and principles for social research. Cambridge, New York etc.: Cambridge University Press; 2007.
65. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.
66. Guo S, Fraser MW. Propensity score analysis. Statistical Methods and applications. Los Angeles, London, New Delhi, Singapore, Washington D.C.: Sage Publications; 2010.
67. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127(8 Pt 2):757-763.
68. Becker SO, Caliendo M. mhbounds – Sensitivity Analysis for Average Treatment Effects. ISA Discussion Paper No 2542 January 2007.

69. DiPrete TA, Gangl M. Assessing bias in the estimation of causal effects: Rosebaum bounds on matcing estimators and instrumental variables estimation with imperfect insturments. Sociological Methodology 2004;34:271-310.
70. Bentzen N, Christiansen T, Pedersen KM. Absenteeism, sickness benefits, and health. Occasional Papers 18/1985, Department of Public Economics, Odense University 1985.
71. Bago dT. Latent class models for use of primary care: evidence from a British panel. Health Econ 2005;14(9):873-892.
72. Deb P, Trivedi PK. The structure of demand for health care: latent class versus two-part models. J Health Econ 2002;21(4):601-625.
73. Deb P, Trivedi PK. Demand for Medical Care by the Elderly: A Finite Mixture Approach. Journal of Applied Econometrics 1997;12(3):313-336.

APPENDIX I

Table A: Q23 What is in your opinion the two most important reasons for the increasing popularity of employer paid health insurance?

(n=1546, those who held HI). Question in HIS dataset ***(Prioritize 1 in the box next to the most important reason and 2 in the box next to the second most important reason)***

Reason	***Most important reason, %***	***Second most important reason %***
Dissatisfaction with the public healthcare system	*9.4%*	***15.3%***
It is a tax free fringe benefit which is free for the employee	***18.5***	***15.3***
It gives access to treatment at private hospitals	***17.4***	***19.4***
Less sickness absence due to quicker treatment	***38.2***	***26.2***
Waiting times in the public healthcare system	***13.7***	***29.2***
Co-payments in the public healthcare system	***0.4***	***0.5***
None of these reasons	***1.2***	***1.2***

Table B: Regression results (OLS and quantile), HIS-dataset:

(STATA, sqreg, 20 rep. for bootstrap). Significance levels indicated by stars, see below table

Dependent: days of absence past 12 months	**OLS**	**Median(q50)**	**q25**	**q90**	**q95**
Characteristics of the employees	Coef	Coef	Coef	Coef	Coef
Health					
Number of chronic diseases	1,1847	0,0730	0,0589	1,6134	1,0966
Self rated health, really good=0					
Good	0,7045	0,3231	0,1305	0,4729	0,5382
Passing	4,3834	1,7886	0,5906	5,3616	10,6774
bad & really bad	19,3065	6,9753	1,7373	50,4948	103,5984
> 6 months illness/consequences of accidents etc0,	-4,2101	-0,2114	-0,1110	-3,2147	-29,4647
Age	1	0,5150	0,2529	1,0640	1,5593
Male (female=0)	-0,1086	-0,0615	-0,0255	-0,1254	-0,1442
No children <15 years	-0,4339	-0,0268	-0,0270	-0,2128	0,3715

Dependent: days of absence past 12 months	**OLS**	**Median(q50)**	**q25**	**q90**	**q95**
Education					
Skilled worker (unskilled =0)	-2,3971	-0,038873	0,0574	-0,6176	-0,4297
Semi-skilled	-1,8539	-0,1461508	0,0644	-1,7370	-0,2376
Junior college	-2,0109	-0,079219	0,0251	-0,4950	-0,9667
College	-2,2229	0,2560131	0,2223	-0,6890	-1,3639
University	-3,1453	0,0580533	0,1566	-1,4317	-2,7082
Mics0,	2,5289	1,0316	0,2180	2,0948	2,3998
Gross income (< 1990,000 =0)					
200-299,000	1,5444	1,2655	0,3368	2,4010	4,6534
300-399,000	1,4172	0,9044	0,1785	1,7509	3,5389
400-499,000	1,7729	0,8410	0,1686	1,2368	4,9405
500-599,000	2,3619	0,7140	0,1868	1,8238	5,2767
600-699,000	2,5474	0,4155	0,1444	0,4075	2,7409
700-799,000	2,3159	1,0506	0,2719	2,8395	8,2658
800,000 and upward	1,4065	0,8790	0,1573	1,4986	5,8676
do not wish to reveal/don't know	2,2731	0,8922	0,1551	0,6796	3,1306
No management responsibility (yes=0)	1,6620	0,7497	0,2897	0,8628	1,9807
Characteristic of place of work					
Public-private:					
Public sector (private sector=0)	-0,1041385	0,3078	0,1026	1,9038	2,2665
Public company	0,7940	-0,2805	-0,2676	0,5643	2,7656
Other/don't know	7,3622	-0,1117	-0,0041	2,9967	14,6841
No0, of employees					
5-9 employess (1-4=0)	0,7077	0,7897	0,2936	0,1915	-1,9361
10-19	0,9145	0,5263	0,2116	0,1805	-2,7286
20 - 49	0,1443	0,7632	0,1658	-0,4379	-2,4810
50 - 99	0,7609	0,8955	0,2428	-0,5811	-3,8138
100 - 249	2,2129	0,7715	0,2095	1,1415	-0,4724
250 - 499	-0,0076	0,3473	0,2014	-0,6675	-2,4015
> 500	2,0478	0,9230	0,3213	0,3213	-1,2552
don't know	6,6469	-0,0431	0,0257	4,3714	2,0472
Healht promotion at work					
Sundhedsordning (yes=0)	0,6216	-0,0976	-0,0316	-0,2235	-1,8571
don't know	-1,4097	0,1509	0,0565	-1,9581	-4,7904

Dependent: days of absence past 12 months	**OLS**	**Median(q50)**	**q25**	**q90**	**q95**
No, of physician contacts past 12 month	0,8892	0,6281	0,2650	1,4878	1,4899
No of A&E and outpatient contacts past 12 month	2,9797	0,9001	0,1777	3,4318	6,4405
Hospitalized (no=0)	18,4364	6,7540	2,9023	67,3891	80,1568
Use of physiotherapy	0,2471	0,1394	0,0232	0,8452	0,6324
Use of Chiropractor	0,4355	0,1412	0,0414	0,1111	0,1652
Healt insurance					
No (yes=1)	1,4607	0,0280	0,0008	0,1882	1,5023
Don't know	3,8258	-0,0414	-0,2890	3,0304	6,8638
Used health insurance past 12 months					
No (yes =0)	*-1,8276*	*-0,4142*	*-0,1824*	*-2,6571*	*-0,8074*
Do not hold health insurane	*-0,6367*	*-0,2756*	*-0,1607*	*-1,6994*	*0,6882*
_cons	3,1728	0,4638	0,0108	9,4607	36,4013
R2	0,17	0.02	0.08	0.25	0.35

Table BB: The full set of regression results for table 7A

Dependent: days of absence past 12 months	Coef.	Std. Err.	z	[95% Conf.	Interval]	Coef.	Std. Err.	z	[95% Conf.	Interval]
Characteristics of the employees	**component1 (0.33)**					**component2 (0.67)**				
Health										
Number of chronic diseases	-0.0298	0.0469	-0.63	-0.1216	0.0621	0.0520	0.0205	2.53	0.0118	0.0923
Self rated health, really good=0										
Good	0.7576	0.2661	2.85	0.2360	1.2792	0.1567	0.0689	2.28	0.0218	0.2917
Passing	0.9642	0.2768	3.48	0.4217	1.5066	0.4033	0.0800	5.04	0.2465	0.5601
bad & really bad	0.9346	0.3126	2.99	0.3220	1.5472	0.8200	0.1065	7.7	0.6112	1.0288
> 6 months illness/consequences of accidents etc.	-0.7665	0.1578	-4.86	-1.0759	-0.4572	0.2624	0.0664	3.95	0.1324	0.3925
Age	-0.0015	0.0064	-0.23	-0.0139	0.0110	-0.0071	0.0024	-2.96	-0.0119	-0.0024
Male (female=0)	-0.2390	0.1296	-1.84	-0.4930	0.0150	-0.0615	0.0462	-1.33	-0.1520	0.0289
Children living at home										
1 child >13 years (no children >13=0)	0.3937	0.1802	2.18	0.0404	0.7469	0.0830	0.0587	1.41	-0.0320	0.1980
2 children > 13 years	0.2824	0.1855	1.52	-0.0811	0.6460	0.1048	0.0575	1.82	-0.0080	0.2176
> 2 children > 13 years of age	0.2229	0.3250	0.69	-0.4142	0.8599	0.1374	0.1027	1.34	-0.0638	0.3387
Education										
Skilled worker (unskilled =0)	-0.4340	0.2289	-1.9	-0.8826	0.0146	-0.1571	0.0887	-1.77	-0.3309	0.0167
Semi-skilled	0.1255	0.2579	0.49	-0.3800	0.6310	-0.1402	0.0987	-1.42	-0.3336	0.0531
Junior college	-0.1181	0.2264	-0.52	-0.5618	0.3257	-0.0247	0.0829	-0.3	-0.1873	0.1378
College	-0.0954	0.2076	-0.46	-0.5022	0.3114	-0.0905	0.0751	-1.2	-0.2378	0.0568
University	-0.2410	0.2386	-1.01	-0.7087	0.2267	0.0746	0.0806	0.93	-0.0834	0.2325
Mics.	-0.6285	0.5639	-1.11	-1.7337	0.4766	-0.2902	0.1460	-1.99	-0.5764	-0.0041
Gross income (< 100.000 =0)										
100-199,000	1.0119	0.6854	1.48	-0.3315	2.3553	0.0488	0.1529	0.32	-0.2508	0.3484
200-299,000	1.2964	0.6656	1.95	-0.0082	2.6010	0.3178	0.1411	2.25	0.0412	0.5944
300-399,000	1.3413	0.6657	2.02	0.0367	2.6460	0.3704	0.1425	2.6	0.0910	0.6497
400-499,000	1.3984	0.6921	2.02	0.0420	2.7548	0.1788	0.1490	1.2	-0.1133	0.4709
500-599,000	1.5558	0.7026	2.21	0.1787	2.9328	0.2908	0.1568	1.85	-0.0166	0.5981
600-699,000	1.2537	0.7638	1.64	-0.2434	2.7507	0.0439	0.1739	0.25	-0.2970	0.3848
700-799,000	1.7708	0.7548	2.35	0.2914	3.2501	-0.1842	0.2100	-0.88	-0.5958	0.2274
800,000 and upward	1.4968	0.7946	1.88	-0.0607	3.0543	-0.1733	0.1911	-0.91	-0.5478	0.2013
do not wish to reveal/don't know	1.3965	0.6763	2.06	0.0709	2.7221	0.0886	0.1529	0.58	-0.2110	0.3882
Form of remuneration										
By the hour (salaried =0)	0.2623	0.2051	1.28	-0.1397	0.6644	-0.1303	0.0865	-1.51	-0.2997	0.0392
Base pay and performance related/piece work	-0.3363	0.2750	-1.22	-0.8753	0.2027	-0.0590	0.0874	-0.67	-0.2304	0.1124
No management responsibility (yes=0)	0.1860	0.1766	1.05	-0.1600	0.5321	0.2512	0.0620	4.05	0.1297	0.3728

Characteristic of place of work										
Public-private:										
Public sector (private sector=0)	-0.1273	0.1576	-0.81	-0.4362	0.1816	0.0950	0.0574	1.66	-0.0175	0.2075
Public company	0.3717	0.3405	1.09	-0.2957	1.0391	-0.3025	0.1990	-1.52	-0.6925	0.0875
Other/don't know	-1.3366	0.6722	-1.99	-2.6541	-0.0190	0.5659	0.1465	3.86	0.2788	0.8530
No. of employees										
5-9 employess (1-4=0)	-0.2014	0.3560	-0.57	-0.8991	0.4963	0.5865	0.1365	4.3	0.3189	0.8541
10-19	-0.1447	0.3462	-0.42	-0.8233	0.5339	0.4216	0.1366	3.09	0.1538	0.6894
20 - 49	-0.4516	0.3306	-1.37	-1.0995	0.1963	0.5091	0.1279	3.98	0.2584	0.7599
50 - 99	0.0097	0.3287	0.03	-0.6346	0.6539	0.4805	0.1300	3.7	0.2257	0.7352
100 - 249	0.1118	0.3267	0.34	-0.5286	0.7521	0.5845	0.1307	4.47	0.3283	0.8408
250 - 499	-0.3330	0.3996	-0.83	-1.1162	0.4501	0.5593	0.1456	3.84	0.2740	0.8447
> 500	-0.2843	0.3240	-0.88	-0.9192	0.3507	0.4796	0.1265	3.79	0.2318	0.7275
don't know	-0.8674	0.5044	-1.72	-1.8561	0.1212	0.6936	0.1685	4.12	0.3633	1.0240
Healht promotion at work										
Sundhedsordning (yes=0)	-0.0402	0.1541	-0.26	-0.3422	0.2619	-0.0128	0.0470	-0.27	-0.1049	0.0794
don't know	0.2674	0.2419	1.11	-0.2068	0.7415	-0.1291	0.0946	-1.36	-0.3146	0.0564
Written policy on sicness absence										
No (yes=0)	-0.0594	0.1607	-0.37	-0.3744	0.2556	-0.1025	0.0583	-1.76	-0.2167	0.0116
Don't know	-0.1688	0.1651	-1.02	-0.4924	0.1549	0.0209	0.0523	0.4	-0.0816	0.1235
Ever called to interview about sickness absence (yes=0)										
No	-0.7687	0.1363	-5.64	-1.0359	-0.5015	-0.5596	0.0555	-10.1	-0.6684	-0.4509
Overall satisfaction with place of work										
Rather satisfied (very =0)	-0.2143	0.1643	-1.3	-0.5364	0.1077	0.0524	0.0553	0.95	-0.0560	0.1608
Satisfied	0.1826	0.1692	1.08	-0.1491	0.5143	0.0508	0.0598	0.85	-0.0664	0.1680
Dissatisfied	0.6059	0.2144	2.83	0.1857	1.0262	0.2435	0.0883	2.76	0.0705	0.4166
Rather and very dissatisfied	0.7661	0.3012	2.54	0.1757	1.3565	0.2935	0.1239	2.37	0.0506	0.5363
Physicially demanding type of work (always=0)										
often	-0.2394	0.2258	-1.06	-0.6819	0.2031	0.0698	0.0991	0.7	-0.1244	0.2640
now and then	-0.0964	0.2219	-0.43	-0.5312	0.3385	0.0571	0.0915	0.62	-0.1223	0.2365
never	-0.2862	0.2254	-1.27	-0.7279	0.1554	0.1730	0.0954	1.81	-0.0140	0.3600
Work pile up if absent (always=0=										
_Often	-0.0608	0.2044	-0.3	-0.4615	0.3399	0.1707	0.0798	2.14	0.0144	0.3270
now and then	-0.1837	0.1872	-0.98	-0.5506	0.1832	-0.0803	0.0750	-1.07	-0.2273	0.0666
never	-0.4281	0.1925	-2.22	-0.8054	-0.0507	-0.1632	0.0750	-2.18	-0.3102	-0.0163
Physically exhausting (always=0)										
_Often	-0.7988	0.2211	-3.61	-1.2321	-0.3654	0.0036	0.1078	0.03	-0.2077	0.2149
now and then	-0.6828	0.2195	-3.11	-1.1131	-0.2525	-0.0278	0.1069	-0.26	-0.2374	0.1818
never	-0.8585	0.2401	-3.58	-1.3291	-0.3879	-0.0358	0.1125	-0.32	-0.2563	0.1846
No. of physician contacts past 12 month	0.3723	0.0808	4.61	0.2140	0.5306	0.1841	0.0288	6.4	0.1277	0.2404

No of A&E and outpatient contacts past 12 month	0.1937	0.1127	1.72	-0.0271	0.4146	0.2902	0.0408	7.12	0.2103	0.3701
Hospitalized (no=0)	1.0493	0.1364	7.69	0.7820	1.3167	0.4963	0.0772	6.43	0.3451	0.6476
Used health insurance past 12 months										
No (yes =0)	-0.4160	0.2060	-2.02	-0.8196	-0.0123	-0.0898	0.0734	-1.22	-0.2336	0.0540
Do not hold health insurance	-0.1761	0.1954	-0.9	-0.5591	0.2069	-0.2345	0.0752	-3.12	-0.3819	-0.0870
_cons	2.1059	0.8808	2.39	0.3796	3.8321	0.5492	0.2728	2.01	0.0145	1.0839

2 component Negative Binomial-1	regression	Number of obs	=	3997
		Wald chi2(124)	=	1829.75
Log likelihood = -9668.8052		Prob > chi2	=	0

/imlogitpi1	-0.705	0.102	-6.93	0	-0.90451	-0.50545
/lndelta1	3.85734	0.112	34.52	0	3.638305	4.076369
/lndelta2	0.49818	0.111	4.47	0	0.279805	0.716556
delta1	47.3391	5.29			38.02734	58.93111
delta2	1.64572	0.183			1.322871	2.04737
pi1	0.33071	0.023			0.288125	0.37626
pi2	0.66929	0.023			0.62374	0.711876

Table BBB: The full set of regression results for table 7A

	Component 1 (0.33)					**Component2 (0.67)**				
Dependent: days of absence past 12 months	**Coef.**	**Std. Err.**	**P>|z|**	**[95% Conf.**	**Interval]**	**Coef.**	**Std. Err.**	**P>|z|**	**[95% Conf.**	**Interval]**
Characteristics of the employees										
Health										
Number of chronic diseases	-0.0329	0.0469	0.483	-0.1248	0.0590	0.0522	0.0200	0.009	0.0130	0.0914
Self rated health, really good=0										
Good	0.7300	0.2646	0.006	0.2113	1.2487	0.1529	0.0683	0.025	0.0191	0.2866
Passing	0.9600	0.2745	0	0.4219	1.4981	0.3974	0.0795	0	0.2415	0.5532
bad & really bad	0.9064	0.3140	0.004	0.2909	1.5219	0.8284	0.1057	0	0.6212	1.0356
> 6 months illness/consequences of accidents etc.	-0.8392	0.1553	0	-1.1437	-0.5347	0.2767	0.0645	0	0.1503	0.4031
Age	-0.0020	0.0063	0.755	-0.0142	0.0103	-0.0070	0.0024	0.003	-0.0118	-0.0023
Male (female=0)	-0.2444	0.1306	0.061	-0.5003	0.0116	-0.0595	0.0451	0.187	-0.1479	0.0289
Children living at home										
1 child >13 years (no children >13=0)	0.3817	0.1782	0.032	0.0325	0.7310	0.0906	0.0576	0.116	-0.0223	0.2036
2 children > 13 years	0.2711	0.1906	0.155	-0.1024	0.6447	0.1098	0.0572	0.055	-0.0024	0.2220
> 2 children > 13 years of age	0.2575	0.3249	0.428	-0.3793	0.8944	0.1448	0.1021	0.156	-0.0553	0.3450
Education										
Skilled worker (unskilled =0)	-0.3826	0.2290	0.095	-0.8313	0.0662	-0.1630	0.0868	0.06	-0.3331	0.0070
Semi-skilled	0.1503	0.2616	0.566	-0.3625	0.6631	-0.1418	0.0993	0.153	-0.3364	0.0527
Junior college	-0.0954	0.2287	0.677	-0.5437	0.3529	-0.0235	0.0826	0.776	-0.1853	0.1383
College	-0.0750	0.2091	0.72	-0.4849	0.3349	-0.0926	0.0742	0.212	-0.2381	0.0529
University	-0.2167	0.2422	0.371	-0.6914	0.2581	0.0722	0.0798	0.366	-0.0842	0.2286
Mics.	-0.5360	0.5416	0.322	-1.5974	0.5254	-0.3061	0.1462	0.036	-0.5926	-0.0196
Gross income (< 100.000 =0)										
100-199,000	1.0069	0.6988	0.15	-0.3627	2.3765	0.0516	0.1526	0.735	-0.2475	0.3507
200-299,000	1.3108	0.6785	0.053	-0.0191	2.6406	0.3161	0.1408	0.025	0.0402	0.5920
300-399,000	1.3289	0.6775	0.05	0.0011	2.6568	0.3780	0.1419	0.008	0.0999	0.6562
400-499,000	1.4243	0.7019	0.042	0.0485	2.8000	0.1823	0.1486	0.22	-0.1089	0.4735
500-599,000	1.6160	0.7139	0.024	0.2169	3.0151	0.2974	0.1564	0.057	-0.0091	0.6039
600-699,000	1.2347	0.7771	0.112	-0.2884	2.7577	0.0539	0.1733	0.756	-0.2857	0.3935
700-799,000	1.7838	0.7655	0.02	0.2834	3.2842	-0.1821	0.2101	0.386	-0.5940	0.2298
800,000 and upward	1.4121	0.8105	0.081	-0.1764	3.0006	-0.1530	0.1920	0.425	-0.5293	0.2232
do not wish to reveal/don't know	1.4107	0.6886	0.041	0.0610	2.7604	0.0973	0.1522	0.523	-0.2011	0.3956

Form of remuneration										
By the hour (salaried =0)	0.2135	0.2046	0.297	-0.1875	0.6146	-0.1201	0.0858	0.162	-0.2882	0.0480
Base pay and performance related/piece work	-0.3428	0.2697	0.204	-0.8715	0.1858	-0.0604	0.0872	0.488	-0.2313	0.1104
No management responsibility (yes=0)	0.2113	0.1761	0.23	-0.1339	0.5564	0.2471	0.0610	0	0.1275	0.3667
Characteristic of place of work										
Public-private:										
Public sector (private sector=0)	-0.1494	0.1568	0.341	-0.4567	0.1580	0.0954	0.0565	0.091	-0.0153	0.2061
Public company	0.3254	0.3285	0.322	-0.3184	0.9693	-0.2738	0.1907	0.151	-0.6476	0.1001
Other/don't know	-1.3923	0.6647	0.036	-2.6950	-0.0896	0.5682	0.1458	0	0.2825	0.8539
No. of employees										
5-9 employess (1-4=0)	-0.1893	0.3668	0.606	-0.9082	0.5296	0.5733	0.1365	0	0.3058	0.8407
10-19	-0.1343	0.3472	0.699	-0.8147	0.5461	0.4173	0.1363	0.002	0.1501	0.6844
20 - 49	-0.4772	0.3382	0.158	-1.1401	0.1856	0.5093	0.1276	0	0.2592	0.7594
50 - 99	-0.0137	0.3341	0.967	-0.6685	0.6410	0.4750	0.1298	0	0.2206	0.7295
100 - 249	0.0794	0.3324	0.811	-0.5721	0.7308	0.5843	0.1301	0	0.3292	0.8394
250 - 499	-0.2483	0.3887	0.523	-1.0101	0.5135	0.5470	0.1415	0	0.2696	0.8244
> 500	-0.2997	0.3320	0.367	-0.9504	0.3510	0.4832	0.1262	0	0.2359	0.7305
don't know	-0.8744	0.5150	0.09	-1.8839	0.1351	0.6857	0.1677	0	0.3571	1.0143
Healht promotion at work										
Sundhedsordning (yes=0)	-0.0383	0.1520	0.801	-0.3362	0.2597	-0.0106	0.0468	0.82	-0.1023	0.0810
don't know	0.2663	0.2479	0.283	-0.2196	0.7522	-0.1361	0.0972	0.161	-0.3267	0.0544
Written policy on sicness absence										
No (yes=0)	-0.0717	0.1581	0.65	-0.3817	0.2382	-0.0958	0.0577	0.097	-0.2089	0.0173
Don't know	-0.1600	0.1665	0.337	-0.4862	0.1663	0.0266	0.0517	0.608	-0.0748	0.1280
Ever called to interview about sickness absence (yes=0)										
No	-0.7807	0.1396	0	-1.0543	-0.5070	-0.5586	0.0551	0	-0.6667	-0.4506
Overall satisfaction with place of work										
Rather satisfied (very =0)	-0.1931	0.1634	0.237	-0.5134	0.1272	0.0573	0.0553	0.3	-0.0510	0.1656
Satisfied	0.2214	0.1700	0.193	-0.1119	0.5547	0.0511	0.0599	0.393	-0.0663	0.1685
Dissatisfied	0.6445	0.2171	0.003	0.2189	1.0700	0.2562	0.0873	0.003	0.0850	0.4273
Rather and very dissatisfied	0.8164	0.2992	0.006	0.2300	1.4028	0.2984	0.1233	0.016	0.0567	0.5401

Physicially demanding type of work (always=0)										
often	-0.2529	0.2274	0.266	-0.6987	0.1928	0.0636	0.0977	0.515	-0.1280	0.2552
now and then	-0.1133	0.2230	0.612	-0.5503	0.3238	0.0510	0.0916	0.578	-0.1285	0.2306
never	-0.3102	0.2260	0.17	-0.7532	0.1327	0.1624	0.0953	0.088	-0.0244	0.3491
Work pile up if absent (always=0=										
_Often	-0.0452	0.2022	0.823	-0.4415	0.3512	0.1652	0.0786	0.036	0.0112	0.3191
now and then	-0.2217	0.1840	0.228	-0.5822	0.1389	-0.0782	0.0739	0.29	-0.2230	0.0667
never	-0.4343	0.1914	0.023	-0.8094	-0.0593	-0.1654	0.0750	0.027	-0.3124	-0.0183
Physically exhausting (always=0)										
_Often	-0.7903	0.2256	0	-1.2325	-0.3481	-0.0043	0.1074	0.968	-0.2147	0.2062
now and then	-0.6404	0.2240	0.004	-1.0795	-0.2014	-0.0364	0.1067	0.733	-0.2455	0.1726
never	-0.8044	0.2424	0.001	-1.2796	-0.3293	-0.0491	0.1116	0.66	-0.2679	0.1697
No. of physician contacts past 12 month	0.3757	0.0808	0	0.2174	0.5341	0.1878	0.0280	0	0.1329	0.2426
No of A&E and outpatient contacts past 12 month	0.2081	0.1080	0.054	-0.0036	0.4198	0.2919	0.0392	0	0.2152	0.3687
Hospitalized (no=0)	1.0686	0.1362	0	0.8018	1.3355	0.4877	0.0741	0	0.3424	0.6329
Healt insurance										
No (yes=1)	-0.0802	0.1549	0.604	-0.3838	0.2233	0.1608	0.0537	0.003	0.0555	0.2660
_cons	1.9615	0.8933	0.028	0.2108	3.7123	0.3094	0.2695	0.251	-0.2188	0.8375

regression	Number of obs	=	3997
	Wald chi2(122)	=	1839.7
	Prob > chi2	=	0

delta1	47.6473	5.35	38.235	59.3761
delta2	1.65312	0.173	1.3467	2.02924
pi1	0.3302	0.021	0.2902	0.37277
pi2	0.6698	0.021	0.6272	0.70976

Table C: Propensity score for health insurance (1=success). HIS data set

Characteristics of the employees	Odds Ratio	Std. Err.	z	P>|z|	95% Conf. Interval	
Health						
Number of chronic diseases	0,9720	0,05	-0,540	0,586	0,8777	1,0766
Self rated health, really good=0						
Good	1,4265	0,1917	2,640	0,008	1,0961	1,8564
Passing	0,9950	0,1643	-0,030	0,976	0,7198	1,3752
bad & really bad	1,2334	0,3455	0,750	0,454	0,7123	2,1358
> 6 months illness/accidents e	0,9636	0,1149	-0,310	0,756	0,7628	1,2173
Age	0,9956	0,0044	-1,010	0,315	0,9869	1,0042
Male (female=0)	1,1654	0,1219	1,460	0,143	0,9493	1,4307
Education						
Skilled worker (unskilled =0)	1,2973	0,2199	1,540	0,125	0,9306	1,8086
Semi-skilled	0,7635	0,1735	-1,190	0,235	0,4891	1,1919
Junior college	1,3855	0,2398	1,880	0,060	0,9869	1,9452
College	0,9568	0,1583	-0,270	0,789	0,6918	1,3233
University	0,8955	0,1649	-0,600	0,549	0,6242	1,2848
Mics,	0,9246	0,4022	-0,180	0,857	0,3941	2,1690
Gross income (< 199,000 =0)						
200-299,000	3,3247	1,1931	3,350	0,001	1,6456	6,7175
300-399,000	5,9450	2,0702	5,120	0,000	3,0043	1,1764
400-499,000	6,6552	2,3809	5,300	0,000	3,3009	1,3418
500-599,000	6,8754	2,5557	5,190	0,000	3,3181	1,4246
600-699,000	1,6261	6,8616	6,610	0,000	7,1113	3,7181
700-799,000	1,0942	5,2413	5,000	0,000	4,2796	2,7979
800,000 and upward	1,3296	5,7396	5,990	0,000	5,7054	3,0986
do not wish to reveal/don't know	4,7700	1,7414	4,280	0,000	2,3322	9,7558
Place of work						
Public-private:						
Public sector (private sector=0)	0,0331	0,0049	-23,150	0,000	0,0248	0,0441
Public company	0,2838	0,0584	-6,120	0,000	0,1896	0,4249
Other/don't know	0,2346	0,0987	-3,450	0,001	0,1029	0,5350
No. of employees						
5-9 employees (1-4=0)	2,8984	0,6552	4,710	0,000	1,8610	4,5141
10-19	5,2432	1,1225	7,740	0,000	3,4463	7,9768
20 - 49	4,9081	0,9682	8,060	0,000	3,3343	7,2248
50 - 99	7,0098	1,5135	9,020	0,000	4,5912	10,7026

Characteristics of the employees	Odds Ratio	Std. Err.	z	P>|z|	95% Conf.	Interval
100 - 249	7,4167	1,5649	9,500	0,000	4,9047	11,2153
250 - 499	11,5280	2,7254	10,340	0,000	7,2529	18,3228
> 500	10,1567	2,0349	11,570	0,000	6858208,0000	15,0416
don't know	0,5178	0,3459	-0,990	0,325	0,1398	1,9181
Healht promotion at work						
Sundhedsordning (yes=0)	0,3153	0,0346	-10,530	0,000	0,2543	0,3908
Union-membership						
LO (not member=0=	0,7114	0,1083	-2,240	0,025	0,5279	0,9587
FTF	1,5877	0,2816	2,610	0,009	1,1215	2,2477
AC	0,5585	0,1251	-2,600	0,009	0,3600	0,8664
Miscel.	1,0471	0,1367	0,350	0,724	0,8108	1,3524
Pseudo R2: 0.37						
N=3265	Documenter in STATA/propensity-hi					

Table D: PSM balancing test, HIS- data set

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
No.of chronic	Unmatched	.92015	.98561	-5.7		-1.63	0.103
	Matched	.92921	.9552	-2.3	60.3	-0.50	0.619
Good health	Unmatched	.61369	.5355	15.9		4.52	0.000
	Matched	.59319	.60036	-1.5	90.8	-0.35	0.730
Passing	Unmatched	.19011	.2413	-12.5		-3.53	0.000
	Matched	.19355	.20012	-1.6	87.2	-0.39	0.696
bad & really bad	Unmatched	.03498	.04362	-4.4		-1.25	0.210
	Matched	.04032	.03196	4.3	3.2	1.06	0.290
> 6 months illness	Unmatched	.75285	.72251	6.9		1.96	0.050
	Matched	.7500	.76912	-4.3	37.0	-1.06	0.291
Male	Unmatched	.43574	.53689	-20.3		-5.81	0.000
	Matched	.43548	.44235	-1.4	93.2	-0.33	0.744
age	Unmatched	43.862	45.356	-13.0		-3.66	0.000
	Matched	44.262	44.111	1.3	89.9	0.32	0.750
Skilled worker	Unmatched	.18099	.12065	16.9		4.93	0.000
	Matched	.18817	.18967	-0.4	97.5	-0.09	0.928
Semi-skilled	Unmatched	.05095	.08167	-12.4		-3.44	0.001
	Matched	.05556	.06243	-2.8	77.6	-0.69	0.491
Junior Coll.	Unmatched	.15589	.10487	15.2		4.43	0.000
	Matched	.14964	.12903	6.1	59.6	1.41	0.160
College	Unmatched	.22281	.32019	-22.0		-6.21	0.000
	Matched	.21953	.20759	2.7	87.7	0.69	0.491

Variable	Sample	Mean Treated	Control	%bias	%reduct |bias|	t-test t	p>|t|
University	Unmatched	.19848	.17865	5.1		1.46	0.146
	Matched	.18548	.181	1.1	77.4	0.27	0.785
Misc.	Unmatched	.01673	.01671	0.0		0.01	0.996
	Matched	.01703	.02897	-9.3	-48267.0	-1.88	0.060
200-299,000	Unmatched	.11711	.23527	-31.4		-8.69	0.000
	Matched	.12455	.11022	3.8	87.9	1.05	0.293
300-399,000	Unmatched	.30570	.32111	-3.3		-0.95	0.343
	Matched	.32258	.33303	-2.3	32.2	-0.53	0.599
400-499,000	Unmatched	.20989	.15638	13.9		4.02	0.000
	Matched	.21147	.22461	-3.4	75.4	-0.75	0.452
500-599,000	Unmatched	.09886	.06357	12.9		3.79	0.000
	Matched	.09857	.07527	8.5	34.0	1.95	0.051
600-699,000	Unmatched	.07529	.0232	24.2		7.40	0.000
	Matched	.05287	.04809	2.2	90.8	0.52	0.606
700-799,000	Unmatched	.03726	.01346	15.2		4.60	0.000
	Matched	.02957	.03286	-2.1	86.2	-0.45	0.656
>= 800,000	Unmatched	.04867	.02227	14.3		4.28	0.000
	Matched	.0448	.05556	-5.8	59.3	-1.16	0.245
Do not reveal	Unmatched	.09734	.10951	-4.0		-1.14	0.256
	Matched	.10573	.1132	-2.5	38.7	-0.56	0.572
Public sector	Unmatched	.07605	.55545	-120.3		-32.28	0.000
	Matched	.07796	.07736	0.1	99.9	0.05	0.958
Public compa.	Unmatched	.04259	.03387	4.5		1.32	0.188
	Matched	.04659	.03913	3.9	14.3	0.87	0.384
do not know	Unmatched	.0076	.02227	-12.1		-3.27	0.001
	Matched	.00896	.00687	1.7	85.7	0.56	0.577
5-9 employees	Unmatched	.05779	.07749	-7.8		-2.21	0.027
	Matched	.06362	.06302	0.2	97.0	0.06	0.954
10-19	Unmatched	.09962	.10812	-2.8		-0.79	0.428
	Matched	.11201	.1132	-0.4	85.9	-0.09	0.929
20-50	Unmatched	.15513	.18283	-7.4		-2.10	0.036
	Matched	.17115	.16189	2.5	66.6	0.59	0.557
50-99	Unmatched	.12395	.12575	-0.5		-0.16	0.876
	Matched	.13351	.13471	-0.4	33.6	-0.08	0.934
100-249	Unmatched	.14981	.12854	6.1		1.77	0.077
	Matched	.16129	.17503	-4.0	35.4	-0.87	0.386
250-499	Unmatched	.10875	.04872	22.4		6.70	0.000
	Matched	.08692	.07198	5.6	75.1	1.30	0.192
>=500	Unmatched	.25475	.15545	24.8		7.24	0.000
	Matched	.21505	.22043	-1.3	94.6	-0.31	0.758
do not now	Unmatched	.00304	.03712	-24.5		-6.37	0.000
	Matched	.00269	.00358	-0.6	97.4	-0.38	0.705
Yes, healt pro	Unmatched	.55475	.79046	-51.9		1476 0000	
	Matched	.61201	.60364	1.8	96.5	0.40	0.686
LO	Unmatched	.25991	.32711	-14.8		-4.19	0.000

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
	Matched	.28136	.32198	-8.9	39.6	-2.09	0.037
FTF	Unmatched	.09832	.12832	-9.5		-2.67	0.008
	Matched	.08871	.05346	11.1	-17.5	3.25	0.001
AC	Unmatched	.04345	.07979	-15.2		-4.19	0.000
	Matched	.04749	.03973	3.2	78.6	0.90	0.369
Misc	Unmatched	.34756	.27671	15.3		4.41	0.000
	Matched	.3414	.38471	-9.4	38.9	-2.13	0.033

Table E: Summary of PSM balancing test, HSI data set

Sample	Pseudo R2	LR chi2	p>chi2
Unmatched	0.368	1594.30	0.000
Matched	0.015	45.98	0.148

Table F: Treatment effect – HIS data set

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Days of sickness absence	Unmatched	7,10	9,54	-2,44	.95750479	-2.55
the past 12 months	ATT	7,09	9,31	-2,22	1,661	-1.34
	ATU	9,37	6,56	-2,82	.	.
	ATE			-2,59	.	.

Table G: Balancing tests, PSR data set

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
	Matched	.73465	.74878	-1.2	91.1	-0.35	0.726
Chronic diseases	Unmatched	.72679	.88579	-13.8		-4.18	0.000
Good health	Unmatched	.59184	.53955	10.6		3.24	0.001
	Matched	.59868	.59381	1.0	90.7	0.26	0.795
Passing	Unmatched	.20145	.2502	-11.7		-3.55	0.000
	Matched	.20541	.20102	1.1	91.0	0.28	0.776
Bad/really bad	Unmatched	.03094	.05287	-11.0		-3.27	0.001
	Matched	.02924	.03509	-2.9	73.3	-0.87	0.386
> 6 month of illness	Unmatched	.82225	.795	6.9		2.11	0.035
	Matched	.82383	.82822	-1.1	83.9	-0.30	0.762
age	Unmatched	41.701	42.622	-8.2		-2.46	0.014
	Matched	41.784	42.053	-2.4	70.8	-0.65	0.513
Male	Unmatched	.605	.47498	26.3		8.05	0.000

Variable	Sample	**Mean** Treated	Control	%bias	**%reduct** |bias|	**t-test** t	p>|t|
	Matched	.59942	.60307	-0.7	97.2	-0.20	0.845
Skulled worker	Unmatched	.15734	.11864	11.2		3.50	0.000
	Matched	.15497	.14766	2.1	81.1	0.53	0.594
semi-skilled	Unmatched	.06583	.08111	-5.9		-1.78	0.076
	Matched	.06652	.07286	-2.4	58.5	-0.65	0.515
junior college	Unmatched	.15537	.11178	12.8		4.01	0.000
	Matched	.15058	.17861	-8.3	35.7	-1.98	0.048
college	Unmatched	.23173	.3301	-22.0		-6.67	0.000
	Matched	.23684	.22466	2.7	87.6	0.76	0.450
university	Unmatched	.2528	.19451	14.0		4.35	0.000
	Matched	.25439	.2539	0.1	99.2	0.03	0.977
Misc.	Unmatched	.02567	.03672	-6.4		-1.91	0.056
	Matched	.02632	.01974	3.8	40.5	1.15	0.251
100-199,000	Unmatched	.02633	.08354	-25.3		-7.34	0.000
	Matched	.02924	.03484	-2.5	90.2	-0.83	0.405
200-299,000	Unmatched	.09151	.17151	-23.8		-7.09	0.000
	Matched	.08333	.09211	-2.6	89.0	-0.81	0.418
300-399,000	Unmatched	.26136	.27845	-3.9		-1.18	0.239
	Matched	.26974	.25755	2.7	28.7	0.72	0.470
400-499,000	Unmatched	.21593	.15456	15.8		4.94	0.000
	Matched	.22368	.2288	-1.3	91.7	-0.32	0.749
500-599,000	Unmatched	.09809	.06699	11.3		3.55	0.000
	Matched	.10526	.09454	3.9	65.5	0.93	0.350
600-699,000	Unmatched	.08558	.03067	23.6		7.68	0.000
	Matched	.07749	.07846	-0.4	98.2	-0.10	0.924
700-799,000	Unmatched	.03292	.01816	9.4		2.98	0.003
	Matched	.03289	.04288	-6.3	32.3	-1.37	0.171
>800,000	Unmatched	.06978	.02663	20.2		6.56	0.000
	Matched	.05775	.06433	-3.1	84.8	-0.72	0.472
Don't reveal	Unmatched	.1106	.11985	-2.9		-0.89	0.376
	Matched	.11184	.09625	4.9	-68.5	1.34	0.182
Hourly pay	Unmatched	.08361	.18483	-30.0		-8.87	0.000
	Matched	.08626	.08845	-0.7	97.8	-0.20	0.839
base pay and piece	Unmatched	.08427	.07829	2.2		0.67	0.500
	Matched	.08114	.0809	0.1	95.9	0.02	0.981
Public sector	Unmatched	.08887	.53511	-		-31.83	0.000
	Matched	.09868	.08821	2.6	97.7	0.94	0.347
Public company	Unmatched	.02567	.02946	-2.3		-0.70	0.482
	Matched	.02778	.03558	-4.8	-106.0	-1.16	0.244
Don't now	Unmatched	.01053	.02179	-8.9		-2.64	0.008
	Matched	.01096	.0173	-5.0	43.7	-1.40	0.161
5-9 employees	Unmatched	.05727	.07587	-7.5		-2.26	0.024
	Matched	.05848	.05117	2.9	60.7	0.84	0.401
10-19	Unmatched	.08361	.11905	-11.8		-3.54	0.000
	Matched	.08553	.08991	-1.5	87.6	-0.41	0.685
20-49	Unmatched	.14812	.16384	-4.3		-1.32	0.186
	Matched	.15351	.15936	-1.6	62.8	-0.42	0.674
50-99	Unmatched	.11126	.13358	-6.8		-2.07	0.038
	Matched	.1155	.12329	-2.4	65.1	-0.63	0.530

Variable	Sample	**Mean** Treated	Control	%bias	**%reduct** |bias|	**t-test** t	p>|t|
100-249	Unmatched	.14417	.11299	9.3		2.90	0.004
	Matched	.14547	.12622	5.8	38.3	1.47	0.142
200-499	Unmatched	.10138	.05811	16.0		5.07	0.000
	Matched	.09064	.10404	-5.0	69.0	-1.18	0.237
> 500	Unmatched	.29427	.20016	21.9		6.84	0.000
	Matched	.28947	.28801	0.3	98.4	0.08	0.933
Don't know	Unmatched	.00461	.03471	-21.8		-6.15	0.000
	Matched	.00439	.00634	-1.4	93.5	-0.70	0.485
Health promotion	Unmatched	.41475	.70541	-61.2		-18.97	0.000
	Matched	.45249	.44834	0.9	98.6	0.22	0.828
Don't know	Unmatched	.04542	.08111	-14.7		-4.37	0.000
	Matched	.04459	.0597	-6.2	57.7	-1.78	0.076
Physically demanding,	Unmatched	.079	.15819	-24.7		-7.32	0.000
	Matched	.08114	.07651	1.4	94.2	0.45	0.653
now and then	Unmatched	.22251	.29298	-16.2		-4.91	0.000
	Matched	.22807	.22027	1.8	88.9	0.49	0.625
never	Unmatched	.63989	.44592	39.7		12.13	0.000
	Matched	.6345	.62719	1.5	96.2	0.40	0.692
Work pile of, often	Unmatched	.13364	.14165	-2.3		-0.71	0.477
	Matched	.13085	.13621	-1.6	33.0	-0.41	0.680
now and then	Unmatched	.34562	.31679	6.1		1.89	0.059
	Matched	.33845	.34917	-2.3	62.8	-0.59	0.555
never	Unmatched	.44437	.41525	5.9		1.81	0.071
	Matched	.45468	.44347	2.3	61.5	0.59	0.556
Physically exhausted,	Unmatched	.10797	.15577	-14.2		-4.27	0.000
	Matched	.10746	.12159	-4.2	70.4	-1.16	0.246
now and then	Unmatched	.41738	.4548	-7.5		-2.31	0.021
	Matched	.42325	.42349	-0.0	99.3	-0.01	0.990
never	Unmatched	.45227	.34948	21.1		6.50	0.000
	Matched	.45102	.43519	3.2	84.6	0.83	0.405