Does the onset of type 1 diabetes in young adults imply increased sickness absence?

Results from a Swedish longitudinal study

Author: Jennifer Ericsson
Master’s thesis 15 credits
April 2008
Supervisor: Carl Hampus Lyttkens
Abstract

Few studies on sickness absence, theoretical as well as empirical, in the field of economics specifically focus on the impact of chronic illness. The purpose of the present study is to fill this theoretical gap by providing a theoretical explanation of sickness absence after the onset of a specific chronic illness, namely type 1 diabetes, but also to empirically examine if sickness absence differs for type 1 diabetics and a control group. The labour supply model of Brown and Session (1996), where health is incorporated in the utility function as an index of sickness, will be used to explain sickness absence in terms of labour-leisure trade-offs. The health capital model developed by Grossman (1972) is used to assess the effect of diabetes on optimal health and time allocation. The main conclusion from the theoretical discussion is that the onset of type 1 diabetes will alter preferences and behaviour. Diabetes is due to an increased mortality rate assumed to lower optimal health stock which will lead to more sick days in a given year in the health capital model. In the labour-leisure model this will increase the risk of sickness absence. Whether absence spells differ between diabetics and non-diabetics is examined with a data set containing type 1 diabetics diagnosed at age 15 to 34 years old, and a control group matched on age, sex and county of residence. These individuals are followed between 1993 and 2005. In the empirical analysis a hurdle model is employed. The empirical analysis show that type 1 diabetics are exposed to a sevenfold excess risk, ceteris paribus, of being absent compared to the control group. However, being stricken with diabetes does not increase the length of the spell.

Key words: Sickness absence, Type 1 diabetes, Hurdle model, The health capital model, The labour-leisure model
Acknowledgements

I owe gratitude to a number of people whose support and assistance made it possible to complete this thesis. First of all I would like to thank Ann-Sofie Brandt at Novo Nordisk for support and many helpful comments. I am very grateful to Katarina Steen Carlsson at Lund University Centre for Health Economics (LUCHE) for providing me the data, but also for excellent guidance and motivation. To Tatiana Nyström at LUCHE, thank you for assistance with the data. Many thanks also to Stefan Leufstedt at Novo Nordisk for enlightening me on what it is like to live as a type 1 diabetic. I am grateful to Novo Nordisk for financial support. Finally I would like to thank my supervisor at the Department of Economics at Lund University, Carl Hampus Lyttkens for help and advice throughout this study.

Lund, April 2008
# TABLE OF CONTENTS

## 1. INTRODUCTION

1.1 BACKGROUND  1
1.2 PREVIOUS RESEARCH  1
1.3 AIM OF THE STUDY  3
1.4 OUTLINE  3

## 2. DIABETES

2.1 WHAT IS TYPE 1 DIABETES?  4
2.2 TREATMENT  4
2.3 COMPLICATIONS  5
2.4 IMPACTS ON EVERYDAY LIFE  6
2.5 ONSET OF DIABETES IN YOUNG ADULTS  6
2.6 COST OF ILLNESS  6

## 3. SOCIAL INSURANCE DURING THE STUDY PERIOD  8

3.1 THE SOCIAL INSURANCE SYSTEM  8
3.2 SICKNESS ABSENCE IN SWEDEN.  10

## 4. THEORETICAL FRAMEWORK  11

4.1 HEALTH AS HUMAN CAPITAL  11
4.1.2 RESULTS FROM THE HEALTH CAPITAL MODEL  13
4.2 THE LABOUR-LEISURE MODEL  14
4.3 LINKING THE HEALTH CAPITAL MODEL AND THE LABOUR-LEISURE MODEL  18
4.4 EMPIRICAL IMPLICATIONS  19

## 5. EMPIRICAL MODEL  21

5.1 THE HURDLE MODEL  21
5.2 A PANEL DATA APPROACH  23

## 6. DATA  25

6.1 STUDY POPULATION  25
6.2 THE DEPENDENT VARIABLES  26
6.3 THE INDEPENDENT VARIABLES  27

## 7. RESULTS  30
1. Introduction

1.1 Background

Sickness absence increased in Sweden in the late 90’s, raising public expenditure drastically in just a few couple of years. Many questions about causes, financing and policy measures of how to decrease absenteeism were raised. Different academic disciplines focus on different explanations of the causes to why an individual absent from work. Anyone could however agree on that illness, although not the only cause, has at least some influence on the decision to report sick from work. From an economic perspective, sickness absence is interesting to study in the case of a chronic illness. In contrast to regular, temporary illness a chronic illness is permanent causing a need for adjustment to the new situation for the inflicted individual. Type 1 diabetes is a chronic illness which occurs when the body is unable to produce insulin, which means that the diabetic needs to inject insulin. Diabetics are exposed to a risk of acute and long term complications as well as a higher mortality rate. In addition, diabetes implies some restrictions on life in terms of a constant need of blood sugar control but also in choices of occupation. This makes type 1 diabetes particularly interesting to study as a diabetic can live an almost normal life, but, as will be explained, the onset of diabetes will lower the health stock. Moreover, in a life time perspective, it will be demonstrated how the health stock would lower even further. In this thesis sickness absence will be analysed from the perspective of how the onset of type 1 diabetes affects the health stock and time allocation which in turn will affect labour-leisure trade-offs. As will be explained, sickness absence would be expected to be higher for type 1 diabetics than for non-diabetics.

1.2 Previous research

Sickness absence has been widely studied in the psychological and organisational field. The main conclusion from these fields is that absenteeism is primarily due to job dissatisfaction or a result of managerial problems. There are relatively few studies on sickness absence in the field of economics. Theoretical work is rare and has mostly elaborated the neo-classical labour-leisure model (Allen, 1981; Chelius, 1981). Other theoretical models include the paper by Chatterij & Tilley (2002) which develops a model based on principal agent theory.

In the existing economic literature on absence few attempts have been made to theoretically include the impact of health status on the decision to report sick. Barmby, Sessions and Treble (1994) model health as an index which was incorporated in a labour-leisure model by Brown and Sessions (1996). Conversely, Kahana and Weiss (1991) model health as a binary variable where the individual will attend if healthy or absent if sick. However, several empirical studies have included the effect of health on sickness absence. Early contributions are Allen (1981) and Paringer (1983) both of which suggest that health is important in explaining absence behaviour. Nonetheless, a large body of work exists examining the effect of health on labour supply in general, workforce participation and earnings (see Currie & Madrian, 1999 for a review).

Relatively few studies specifically examine the impact of diabetes on sickness absence. Moore and Buschbom (1974) present a literature review in addition to their own study. The main result of the studies in the literature review, although diverging, is that diabetics have more sick days. However, the authors discuss the reliability of these results. In their own study, the authors conclude that diabetics are not more absent than the general population. Waclawski (1990) examined sickness absence among insulin treated employees in the UK and found that diabetics and the control group had a similar frequency of absence. However, the diabetics had a greater average length of absence spell. Poole et al (1994) study 91 diabetics and non-diabetics at an engineering factory. Sickness absence was greater for the diabetics both in terms of number of spells and length of spells, but neither was significant.

More recent studies are Skerjanc (2001) who compares sickness absence among 400 diabetic and non-diabetic employees in Ljubljana, Slovenia. The result indicates that diabetes is associated with both longer and more frequent sickness absence spells. Tunceli et al. (2007) study the effect of glycemic control on absenteeism. This study included 233 employed patients diagnosed with diabetes in south-east Michigan in the US. The result of the study shows that poor glycemic control, in some cases, was associated with increased absenteeism. Kivimäki et al (2007) examine if the excess risk of sickness absence was accounted for by diabetes itself or for coexisting conditions. The study cohort consisted of 33 148 Finnish public sector employees. The results show that diabetes was associated with an increased risk
of absenteeism. Of this excess risk 55 percent was due to co-existing non-cardiovascular diseases, e.g. depression. Cardiovascular complications contributed with a 7 percent increased risk of absence.

1.3 Aim of the study

In the economics field there is a clear lack of studies examining the effects of a chronic illness on sickness absence. Moreover, few empirical studies investigate whether diabetics are more prone to be absent. This study aims at filling this gap. The aim of the present study is thus to theoretically examine the effects of diabetes onset but also to empirically test if absence spells differ between diabetics and non-diabetics. The labour supply model of Brown and Session (1996), where health is incorporated in the utility function as an index of sickness, will be used to explain sickness absence in terms of labour-leisure trade-offs. The health capital model developed by Grossman (1972) is used to assess the effect of diabetes on health and time allocation. The study thus aims at answering the following questions;

What are the theoretical implications of type 1 diabetes in explaining sickness absence? 
Does sickness absence spells differ between diabetics and the control group?

Whether absence spells differ between diabetics and non-diabetics is examined with a data set containing type 1 diabetics diagnosed at age 15 to 34 years old, and a control group matched on age, sex and county of residence. These individuals are followed between 1993 and 2005.

1.4 Outline

An introduction to diabetes and the sickness insurance will be given in chapter two and three, respectively. In chapter four, the theoretical implications of diabetes onset will be developed. Chapter five continues with a description of the empirical model and the data the present study is based on will be presented in chapter six. The result of the study will be presented and analysed in chapter seven. This study will conclude with a discussion in chapter eight.
2. Diabetes

2.1 What is type 1 diabetes?

Diabetes is a chronic illness that occurs when the blood glucose, or blood sugar, is too high. There are several types of diabetes. The most common one is type 2 diabetes which means that pancreas is able to produce insulin, a hormone that helps glucose to enter cells to give them energy, but not sufficiently. Type 1 diabetes means that pancreas produces little or no insulin. In Sweden, about 300 000 people are afflicted with type 2 diabetes and about 50 000 with type 1 (The Swedish Diabetes Association). Type 1 diabetes is also known as juvenile diabetes since the majority is diagnosed when they are less than 14 years old. About 300 people are diagnosed yearly when they are between 15-34 years old but it is also possible, but rare, to be diagnosed when even older. The exact causes of type 1 diabetes are not known. Diabetes is partly hereditary, partly caused by environmental changes. Supposedly, an immune system that attacks and destroys insulin producing cells is inherited, a process initiated by environmental changes. It is however not known what environmental changes initiate this process.

Carbohydrates are the most important source of energy for the body. The carbohydrates break down to glucose, or sugar, in the intestines from where it is taken up by the blood. The glucose is transported by the blood to organs where it is used as a source of energy for the cells. Insulin is a hormone needed to make glucose enter the cells. Without insulin too much glucose stays in the blood where the body is unable to use it for energy.

2.2 Treatment

Besides medication diabetes treatment is to a large extent built on self medication such as maintaining a good diet and exercising. Insulin treatment aims at keeping a tight blood sugar control. Since there is yet no cure to diabetes, emphasis is on managing and preventing short term and long term complications. Self medication is a vital part of diabetes treatment and it is therefore crucial that the diabetic has great knowledge about how to handle and live with the disease. Generally, the diabetic needs to inject insulin four to five times per day. Insulin injections are adjusted to each individual but a common procedure is to inject one long-acting
insulin shot once per day and a short-acting insulin shot before every meal. Keeping a good diet is essential for a diabetic. A good diet consists of much fruit and vegetables and fewer calories as well as fat, which is essentially the same as for non-diabetics. Dietary guidelines for type 1 diabetics are instead focused on how food affects blood glucose and insulin treatment. It is however important to spread the meals as evenly as possible during the day. Moreover, to exercise regularly is also important to a diabetic. In addition to general health improving effects, exercise improves insulin sensitivity which is important since high insulin sensitivity improves the effect of insulin treatment.

2.3 Complications

Diabetics are exposed to a risk of several complications. Acute complications are hypoglycaemia, low blood sugar, and hyperglycaemia, high blood sugar. When hypoglycaemia develops, cells are not getting enough glucose. Eventually, the low level of blood sugar leads to confusion, loss of consciousness, coma and death. Hyperglycaemia might lead to ketoacidosis, the accumulation of ketones in the blood when the body uses fat instead of glucose for energy. Ketones make the blood acidic and slow down all body functions which, like hypoglycaemia, can lead to coma and death. Long term complications might arise since diabetes affects the arteries which increase the risk of heart attacks, stroke and gangrene in the legs. Diabetes also damages the small blood vessels which could cause impaired vision and blindness as well as kidney disease. In addition, diabetes could cause nerve damage and impaired heart- and intestine function. Complications might arise after several years of diabetes. A tight blood sugar control might prevent the emergence of complications. Wibell et al. (2001) show that diabetes in young adults (15-34) is associated with increased mortality. Compared to the control group, diabetics had a threefold larger mortality rate. The excess mortality was not primarily due to diabetes as such but rather to environmental factors such as alcohol and drug abuse as well as mental instability in combination with diabetes. A more recent study show a two-fold increases in excess mortality in individuals with type 1 diabetes onset as young adults (Waernbaum et al., 2006).
2.4 Impacts on everyday life

Diabetes affects life in several ways. The diabetic needs to keep close control of blood sugar levels. Food and all kinds of emotions such as stress or excitement, change blood sugar levels. Diabetes also implies several restrictions on life. The risk of hypoglycaemia and the dependence on insulin injections and food affect choice of occupation. It is for example not possible to work in the police force, become a pilot or, in most cases, at sea. Furthermore, there are restrictions regarding driver’s license and military service.

2.5 Onset of diabetes in young adults

There are differences between childhood and adulthood onset of type 1 diabetes. A recent comparison among several European countries indicates differences in incidence where the incidence in childhood diabetes is higher than in adulthood onset (Kyvik et al., 2004). Another result from this study is that there is a male predominance in type 1 diabetes with onset in adulthood compared to onset in childhood with small differences. An example of an environmental factor associated with the onset of diabetes is stressful events. Littorin et al. (2001) show that hereditary factors seems to be dominated over stressful events in the onset of type 1 diabetes as a young adult. Mortality rates also seem to differ depending on the age of diabetes onset such that a younger age means higher mortality (Borch-Johnsen, 1986). Jonsson et al (2000) investigated health care costs the first ten years after diabetes diagnosis as a young adult. The authors estimate significantly higher health care costs compared to the control group. Interestingly, this is even before complications might have developed.

2.6 Cost of illness

Diabetes incurs large costs on society (The National Board of Health and Welfare). Costs are due to direct treatment costs, cost of pharmaceuticals and assistance as well as indirect costs due to lost production. There is no recent cost of illness study due to type 1 diabetes only in Sweden. However, Henriksson and Jönsson (1998) estimated both direct and indirect costs of type 1 and type 2 diabetes in 1994. The total cost of diabetes was estimated to 5.7 million SEK\(^1\). Direct costs constitute about 43 percent of the total costs while indirect costs thus

\(^1\) 1 US$ = 7.5 SEK in 1994
amount to 57 percent. The indirect costs consisted of production losses due to morbidity and premature mortality and are the dominant costs. Moreover, Norlund et al (2001) estimate the cost of illness including comorbidity using data from a Swedish municipality in 1993. The direct and indirect costs of diabetes are estimated to be 2.5 times higher than earlier estimates, ie. than the estimate in Henriksson and Jönsson (1998). The main conclusion is thus that it is necessary to consider comorbidity as well.
3. Social insurance during the study period

3.1 The social insurance system

All residents in Sweden are entitled to sickness insurance for an unlimited period, provided an annual income of at least 24 per cent of a “basic amount”\(^2\), if work capacity is reduced by at least one quarter. Not all income is however reimbursed, as 7.5 times the basic amount serves as an upper ceiling of an income covered by sickness insurance. Sickness insurance is financed through payroll taxes.

The first day of illness the individual reports sick to the employer. During the first six days, the decision as to whether the individual is able to work or not is based on the individual’s own perception of his ability to work. For continued compensation a doctor’s certificate will however be needed. As The Social Insurance Agency decides whether the individual is entitled to sickness insurance a doctor’s certificate is however no guarantee of reimbursement. It is the ability to work that decides whether there is any ground for sickness insurance, not the illness or symptoms as such. The first day of a sickness spell will not be reimbursed. During the first 14 days, the employer is responsible for sickness insurance which is 80 percent of the individual's salary during this period. After these 14 days, the individual will receive sickness insurance from The Social Insurance Agency. The individual will then be reimbursed with about 80 percent, according the prevailing reimbursement rate in table 1, of their annual salary up to 7.5 times the basic amount.

The reimbursement rate changed several times during the study period. Table 1 summarises these, and some other, changes. A waiting day was introduced the 1 of April 1993, in order to weaken incentives to absent. As a step towards reducing the increasing costs of absence during the beginning of the 21\(^{st}\) century the income which sickness insurance is based upon is multiplied by the factor 0.97 before calculating sick pay. At the same time the reimbursement rate is reduced. As stated above, the employer is responsible for the first 14 days of an absence spell. In 1997 this was changed to the first 28 days, and then changed back to 14 days in 1998. This changed yet again in 2003 to include the first 21 days which was changed back in 2004.

\(^2\) The basic amount is determined by the government and is closely connected to the consumer price index. In 2005 the basic amount was 39 400 SEK.
Refers to calendar days, not working days.

This reform is ongoing since January 1992.

From August 2003 and onwards the income to which sick pay is based on is multiplied by the factor 0.97.

Waiting day

---|---|---|---|---|---|---|---|---|---|
1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
2-14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
15-21 | 80 | 80 | 80 | 75 | 0 | 0 | 80 | 0 | 80 |
22-28 | 80 | 80 | 80 | 75 | 0 | 0 | 80 | 77.6 | 80 |
29-90 | 80 | 80 | 80 | 75 | 75 | 80 | 80 | 77.6 | 80 |
91-365 | 90 | 80 | 80 | 75 | 75 | 80 | 80 | 77.6 | 80 |
365- | 90 | 80 | 70 | 75 | 75 | 80 | 80 | 77.6 | 80 |


Table 1 Changes in the reimbursement rate

---

3 Refers to calendar days, not working days.
4 This reform is ongoing since January 1992
5 From August 2003 and onwards the income to which sick pay is based on is multiplied by the factor 0.97
6 Waiting day
In addition to social sickness insurance about 90 percent of all employees are covered with supplementary insurance (Ståhlberg, 2006). Supplementary insurance is negotiated by collective agreements and usually reimburses an additional 10 percent above the 80 percent reimbursed through social insurance. This means that almost all employees in Sweden are reimbursed with 90 percent of the income loss during a sickness spell. Furthermore, employees with incomes above the social insurance ceiling are still reimbursed with 90 percent of their total income with supplementary insurance. Supplementary insurance is thus especially important for high income earners.

3.2 Sickness absence in Sweden.

After a period of high sickness absence in the late 80’s, sickness absence decreased between 1990 and 1997. This was followed by a drastic increase between 1997 and 2003, which caused national expenditures on sickness insurance to increase from 15 billion SEK in 1997 to 44 billion in 2002 (National Institute of Economic Research). During these years, Sweden had the highest rates of absence in Europe. There are also large regional differences in sickness absence. High sickness absence is concentrated to the northern parts of Sweden but also to sparsely populated areas in general (The Swedish Social Insurance Agency). Figure 1 show the percent of the labour force absent between 1993 and 2005.

Figure 1 Percentage absent of the labour force, 20-64 years old  Source: Statistics Sweden
4. Theoretical framework

The theoretical discussion aims at explaining whether sickness absence differs between type 1 diabetics compared to non-diabetics. The health capital model of Grossman (1972) is used to assess how the onset of diabetes affects optimal health stock and available time. A labour supply model where health is included in the utility function, as developed originally by Barmby, Sessions and Treble (1994) and explained in a labour supply framework by Brown and Sessions (1996), is used to describe sickness absence in terms of labour-leisure tradeoffs. Next, it will be assessed how the results from the health capital model affects the labour-leisure choice in the labour supply model and vice versa.

A central terminology in this study is the concept of a health stock, originally used in the health capital model (Grossman, 1972). The health stock of an individual can be defined as current health but does also include longevity. The onset of type 1 diabetes is called a negative health shock since it reduces the health stock of an individual.

4.1 Health as human capital

Health is an important human capital and vital to the explanation of absence behaviour. The health capital model of Grossman (1972) is built on human capital theory, according to which individuals invest in themselves through education, training and health to increase earnings. Grossman shows that the demand for health differs in several aspects as compared to the demand for other commodities. Firstly, it is not medical care per se the individual demands but rather good health. Secondly, individuals do not purchase health from the market but are instead producing it themselves by making health improving efforts. Furthermore, health is regarded both as a consumption commodity and as an investment commodity. As a consumption commodity, health is demanded because it makes you feel better. As an investment commodity health yields monetary, not utility gains, and is demanded as it increases the number of healthy days available to work and earn income. Finally, health is viewed as a durable capital stock, yielding healthy time, and thus last for more than one period.
In the health capital model, individuals derive satisfaction from health and from the consumption of other commodities. Furthermore, individuals inherit an initial stock of health that depreciates over time. The stock of health today depends on previous health investments and the rate of depreciation, such that

\[ H_{i+1} = H_i(1 - \delta_i) + I_i \]  

where \( H_i \) is health in period \( i \), \( I_i \) is gross investment and \( \delta_i \) is the rate of depreciation. The rate of depreciation is assumed to be exogenous and increases with the age of the individual. Individuals invest in health capital by devoting time and medical care to the production of health. Individuals are therefore able to “choose” their health status as well as their length of life. An important feature of the model is that health is endogenous and depends on the resources allocated to its production in addition to the initial inherited stock of health. The uses of healthy time consist of work time, time spent producing health and time spent producing other commodities. These activities are mutually exclusive and sums up to total time. Time spent being sick is subtracted from total time and is assumed to be lost time. Leisure consists of time spent producing health and time spent producing other commodities.

A marginal efficiency of capital curve (MEC) can be drawn as in figure 2 (Grossman, 1972). The MEC curve shows the relationship between the stock of health and the marginal efficiency of health capital, \( \gamma_i \). The supply curve is assumed to be constant, and equal to the market rate of interest, \( r \), the marginal cost of gross investment in health \( \pi_{i-1} \), and the rate of depreciation, \( \delta_i \). The optimal health stock in period \( i \) is \( H^* \).

![Figure 2](image-url)  

*Figure 2* The demand curve for health (Grossman, 1972)
In the health capital model, health is also analysed in a life cycle perspective by examining the effect of the depreciation rate on the demand for and investments in health (Grossman, 1972). The depreciation rate is assumed to grow continuously with age after some point in the life cycle and the health stock will thus decline over time. The effect of declining health on the demand for health producing time is unclear. However, Grossman concludes that if the elasticity of the MEC curve were less than one, gross investment and the depreciation rate would be positively correlated while gross investment and the health stock would be negatively correlated over the life cycle. This means that individuals would desire to offset part of the reduction in the health stock, caused by the increased depreciation rate, by increasing gross investment and hence the demand for health producing time.

4.1.2 Results from the health capital model

In the health capital framework, a negative health shock, such as the onset of type 1 diabetes, has several effects. The first effect is an increase in the rate of depreciation due to the fact that type 1 diabetes increases mortality. The higher depreciation rate increases the cost of holding capital which results in less health capital being held. In figure 2, this implies a shift in the supply curve from $S^*$ to $S'$ resulting in a fall in the optimal health stock. The health stock and the number of healthy days are positively correlated (Grossman, 1972). Therefore, as the stock of health increases the number of healthy days also increases, although at a diminishing rate. Conversely, a decline in the optimal health stock will increase sick days. Hence, the onset of type 1 diabetes will result in a lower optimal health stock as well as more time being lost to illness. As sick time is regarded as lost time, time being sick will result in less time over for work and leisure.

The health capital model also predicts the increased depreciation rate to increase demand for gross investment in health, given elasticity less than one of the MEC curve (Grossman, 1972). Hence, the demand for health producing time would increase. The effect of an increased demand on health producing time is a decline in time available for other uses, that is, working time and time spent producing other commodities as these activities are assumed to be mutually exclusive. Thus, in the health capital model, it is not possible to maintain health while working. As a result, the marginal utility of leisure will increase relative to the marginal utility of work.
Moreover, the onset of diabetes causes a sudden fall in the health stock. The health stock could initially fall below the new optimal health stock, leading to even more days lost to illness. For a type 1 diabetic, this could be the cause of the often acute development of the illness as well as time needed for changing and adapting a life-style in accordance with the illness. In addition, it could also take some time to find the appropriate treatment. This would lead to more sick days immediately after the illness strikes.

A central assumption in the health capital model is that health is not a determinant of the wage rate (Grossman, 1972). Instead, health impacts income primarily through the number of sick days. Moreover, educated individuals are assumed to be more efficient producers of health capital. A higher education would then shift the MEC curve in figure 2 to the right, given a constant wage rate and marginal product of health. As the cost of capital is independent of education there would be no shift in the supply curve. Hence, more educated would demand a larger optimal health stock.

4.2 The labour-leisure model

Sickness absence can be analysed within the traditional neo-classical labour-leisure choice model. In this model, the individual receives satisfaction from the consumption of goods and from leisure. The model is based on the trade off between leisure, which yields no money, and work time, which yields the money necessary to the desired consumption of goods. An individual's labour supply schedule represents the amount of work the individual is willing to offer at different wages. In this schedule the utility of different combinations of work, income and leisure is shown. By maximising the individual's utility subject to alternative wage rates, the optimal combinations of labour and leisure are derived. In the long run, it is predicted that the individual's desired amount of work effort will prevail as actual hours of work. In the short run, on the other hand, a predetermined contract might state a number of hours of work that exceeds the desired work effort. Absence occurs when the individual brings actual hours of work into equality with desired hours (Chelius, 1981). This study however aims at extending this traditional labour-leisure model by including the effect of health status on labour supply. The presentation of the theoretical model follows the model developed in Brown and Sessions (1996).
The point of departure in Brown and Sessions (1996) is that individuals are risk neutral utility maximisers endowed with a stock of time, $T$, which they allocate between work and leisure. Utility is, as in the traditional labour-leisure model, an increasing function of consumption and leisure. Health is formally incorporated in the utility function as an index of sickness, $\sigma$. This index is assumed to be a random variable with a probability function $f(\sigma)$, and randomly distributed over the interval [0,1]. In other words, the individual is exposed to a health shock which generates a different value of $\sigma$, that is, a new level of sickness, and hence a new utility maximisation problem. The parameter $\sigma$ represents the individual’s general level of health with higher levels of $\sigma$ indicating higher levels of sickness. The weight placed on leisure relative to consumption is interpreted as the index of sickness such that individuals value leisure more as $\sigma$ approaches one. This is a result of more time needed for recuperation or because it becomes onerous to work at higher levels of sickness. The utility function is

$$U = U(x, l, \sigma)$$  \hspace{1cm} (2)

where $x$ represents the individual’s consumption of goods and services, $l$ represents leisure time and $\sigma$ represents the level of sickness. Furthermore, individuals are assumed to have Cobb-Douglas preferences. The utility function can thus be written as

$$U = x^{(1-\sigma)}l^{\sigma}$$  \hspace{1cm} (3)

Sickness absence becomes possible when there are discontinuities in the budget set, that is, when the individual is not able to freely select his utility maximising set of leisure hours. An example, where the budget set consists of only two points, $E^c$ and $E^s$, will illustrate how sickness absence becomes possible. This is illustrated below in figure 3.
The budget set outlined in this example gives the individual two opportunities. At point $E_c = \{x_c, l_c\}$ the individual attends work with leisure specified as $l_c = T - h$, where $h$ is the hours of work in the employment contract. Sickness absence is described by point $E_s = \{x_s, T\}$. Obviously, total time is devoted to leisure at this point. The individual maximises the utility function

$$\max_U = x^{(1-\sigma)} l^\sigma$$

When maximising utility the individual is bound by two constraints. The first one concerns available time and the second one income. The budget set in this presentation consists of two points where each point is subject to the two constraints. The constraints for attending work, point $E_c$, are

- Time constraint: $l = T - h$
- Income constraint: $x = w$

The constraints for sickness absence, point $E_s$, are

- Time constraint: $l = T$
- Income constraint: $x = s$

where $s$ is sick pay. It is assumed that $s < w$. The utility function in (4) clearly shows that different values of $\sigma$ yield different looking indifference curves, where higher values of $\sigma$ lead to a relatively steeper curve and vice versa. An increase in $\sigma$ increases the marginal rate of
substitution between consumption and leisure meaning the individual places a greater value on leisure.

The value of $\sigma$, the level of sickness, will determine whether the individual’s utility is maximised at point $E^c$ or $E^s$, i.e. whether she will attend work or report sick. This decision will depend on a “reservation” level of sickness, $\sigma^*$, which serves as a reference point for the individual. The reservation level of sickness is defined as the level of sickness where the individual is indifferent between attendance and absence. More formally, it is defined as where the utility of attending equals the utility of being absent

$$w^{(1-\sigma^*)}(T - h)\sigma^* = s^{(1-\sigma^*)}T^{\sigma^*}$$ (7)

The reservation level of sickness, $\sigma^*$, is obtained by solving (7)

$$\sigma^* = \frac{\ln(w/s)}{\ln(w/s) + \ln[T/(T-h)]}$$ (8)

In figure 3, the indifference curve $U(\sigma^*)$ represents the reservation level of sickness. When exposed to a health shock, a new value of $\sigma$ is generated determining the slope of the new indifference curve and whether the individual will absent or attend work. If $\sigma > \sigma^*$ the individual maximises utility by being absent since the slope of the indifference curve will be steeper than that of $U(\sigma^*)$, for example $U(\sigma_2)$. If, on the other hand, $\sigma < \sigma^*$ the slope of the indifference curve is flatter, for example $U(\sigma_1)$, then the individual maximises utility by attending work. It is thus the value of $\sigma$ generated by the health shock that decides if the utility is maximised subject to (5) or (6).

Clearly, the reservation level of sickness in equation (8) depends on the wage, sick pay and contracted hours of work. An increase in the wage rate will increase the reservation level and thus decrease expected absence while an increase in sick pay will decrease the reservation level and increase expected absence. Moreover, if contractual work hours, $h$, were to decrease, the reservation level would increase.

In theory, the marginal utility of consumption relative to leisure could increase as health decreases (Currie & Madrian, 1999). This stems from an increased need of commodities such as health treatments, medicines or assistance in everyday life. However, when it comes to
diabetes, these commodities are subsidised and it is therefore reasonable to assume this effect to be insignificant\(^7\). The labour-leisure model outlined here therefore applies.

4.3 \textit{Linking the health capital model and the labour-leisure model}

The health capital model suggests that the onset of diabetes leads to a decrease in optimal health stock will decrease. In the labour-leisure model this generates a permanent increase in the value of \(\sigma\), the level of sickness. The onset of diabetes thus generates a new maximisation problem and a new indifference curve. It is not likely that the onset of diabetes will generate a level of sickness that exceeds the reservation level of sickness, since this would mean that the individual withdraws from the labour force. Hence the new level of sickness caused by the onset of diabetes generates a steeper indifference curve that lies closer to the reservation level in figure 3. This will increase the probability of another health shock generating a value of \(\sigma\) higher than the reservation level. Consequently, the probability of reporting sick increases after diabetes onset.

A second effect arises due to an increased demand for health maintenance. In the health capital model this implies less time for work and time spent producing other commodities as these activities are mutually exclusive. As leisure consists of time spent producing health and other commodities, the marginal utility of leisure relative to work would increase. Increased demand for health maintenance time should be interpreted as increased marginal utility of leisure also in the labour-leisure model. This is due to the fact that only a certain amount of health maintenance activities could be assumed to be allowed to be performed during work time. Furthermore, the health capital model predicts more sick days due to the lower optimal health stock. This will “steal” time from work and leisure as sick time is regarded lost time (Grossman, 1972). The labour-leisure model is however not as restrictive as it would be possible to be sick while e.g. at work.\(^8\) Obviously, this is because sickness is interpreted as a continuous index, where the question instead is weather one’s sickness level is above or below the reservation level. However, more sick days would be interpreted as a higher level of sickness in the labour-leisure model. A higher level of sickness implies a steeper indifference curve and hence a higher marginal utility of leisure relative to work. The general

\(^7\) In Sweden, insulin is free of charge for the diabetic.

\(^8\) The labour supply model and the health capital model are fundamentally different regarding the view on time spent being sick. In the health capital model, time spent being sick is regarded as lost time while in the labour-leisure model time spent being sick is leisure if the level of sickness is higher than the reservation level.
conclusion from the health capital model and the labour-leisure model on the allocation of
time is that the onset of type 1 diabetes will increase marginal utility of leisure relative to
work.

The health capital model suggests that health does not determine the wage rate (Grossman,
1972). In the labour-leisure framework, on the other hand, this is not as obvious. A negative
health shock might lower productivity and hence the wage (Currie & Madrian, 1999). It is
clear from equation (8) that a lower wage rate will decrease the reservation level of sickness
and increase expected absence. However, it would be reasonable to assume that diabetes
should not affect productivity, at least as long as the diabetic does not have any complications.
On the other hand, an employer with knowledge about the health shock could, due to
prejudices, perceive the diabetic as less productive. Hence, the wage could decrease
regardless of whether the diabetic is less productive or not. If the wage rate were to decrease,
this would result in the MEC curve, in figure 2, shifting inwards in the health capital model.
This would decrease optimal health stock furthermore. Consequently, over the life cycle, a
lower wage would also contribute to more sick days for diabetics.

The health capital model predicts higher educated individuals to demand a larger optimal
health stock (Grossman, 1972). Higher educated diabetics (and non-diabetics) would then
have a larger health stock and consequently fewer days lost to illness. In the labour-leisure
framework, this would be interpreted as higher educated individuals having a lower value of $\sigma$
and hence a flatter indifference curve.

The fact that complications might arise in time would increase the depreciation rate further
and thus decrease optimal health stock yet again. More sick days would then be predicted for
diabetics. Deteriorating health gives a permanent increase in the level of sickness in the
labour-leisure model. In time, this would thus increase the level of sickness and steepen the
indifference curve, increasing the probability of absence further.

4.4 Empirical implications

The main result from the health capital model is that the onset of diabetes lowers optimal
health stock due to increased mortality. This implies more sick days for diabetics. In the
labour-leisure model the lower optimal health stock increases the probability of reporting sick.
Moreover, both the health capital model and the labour-leisure model suggest the marginal
utility of leisure to increase relative to work due to increased demand for health maintenance
time. The higher marginal utility of leisure to work increases preferences for reporting sick.
Altogether, the models predict both an increased probability of sickness absence and longer
spells for type 1 diabetics compared to non-diabetics. In a life-time perspective, the health
capital model predicts an even lower optimal health stock if complications arise which would
increase probability and length of an absence spell further.

These models suggest several variables to be included in the empirical analysis. From the
labour supply model, the wage rate, contractual work hours and sick pay would influence the
number of days absent from work. The level of sickness is measured as having type 1 diabetes
or not, where suffering from diabetes increases the probability of being absent. In addition,
the health capital model suggests that a higher education implies fewer days absent. It does
however not indicate whether education would differ between diabetics and non-diabetics.
The health capital model also suggests that age should be included, which is also a proxy for
health. A lower wage could also be expected for diabetics, which is why this should also be
incorporated in the empirical model. Time since diagnosis is also included as days absent and
the probability of absence will also increase with time.
5. Empirical model

The dependent variable in this analysis, the number of days for which the individual received sickness insurance, is a count; a non-negative integer value. It is therefore not possible to use traditional linear models as failing to account for the non-negativity would generate biased estimates, impossible predictions and invalid inference (Winkelmann, 1997). Count data models, such as the Poisson and the Negative Binomial model have been developed in order to account for this. Moreover, a large fraction of the individuals in this data set does not have any days absent at all, that is, all zero values. In fact, only absence spells longer than 14 days are analysed. Hence, the dependent variable is truncated at zero which also calls for special statistical properties. The hurdle count data model originally proposed by Mullahy (1986) is therefore employed in order to address the excess zeros\(^9\). The hurdle model consists of two processes; the first one governs the probability of passing the hurdle and the second one the length of the absence spell once the hurdle has been passed.

5.1 The hurdle model

A general probability distribution of the hurdle model is given by first considering the probability of a zero outcome (Winkelmann, 1997),

\[
P(Y = 0) = f_1(0)
\]  

and the probability of crossing the hurdle,

\[
P(Y > 0) = 1 - f_1(0)
\]

Assume that \(f_1\) and \(f_2\) are any probability distributions for non-negative integer values, such as the Poisson or the negative binomial distribution. The hurdle part is governed by \(f_1\) while \(f_2\) governs the process once the hurdle has been crossed. The method of deriving the conditional

---

\(^9\) Another approach would be to employ a Tobit model which also accounts for a large proportion of zeros. This has been done in e.g. Leigh (1986) and Dunn & Youngblood (1986). The main advantage of using a hurdle model is that this model allows for different variables (and different signs) to enter the equations. The Tobit model on the other hand, is a single equation model assuming both the participation and the conditional equation to be a result of the same process.
distribution, that is the process once the hurdle has been crossed, is based on the usual conditional probability rule.

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} \]  

(11)

The probability distribution of the hurdle model thus consists of two parts. The first one is the probability of a zero outcome and the second one is the conditional probability,

\[ P(Y = 0) = f_1(0) \]  

(12)

\[ P(Y = y|Y > 0) = f_2(y) \frac{1 - f_1(0)}{1 - f_2(0)}, \quad y = 1, 2, ... \]  

(13)

The hurdle model can be specified in various ways by choosing different probability distributions \( f_1 \) and \( f_2 \). The chosen probability distributions need not, but could be the same. It is easy to show that the two parts can be estimated separately by considering the log-likelihood function of the hurdle model

\[
\log(L) = \sum_{y=0} \log[1 - P_1(y > 0|x)] + \sum_{y>0} \left\{ \log[P_1(y > 0|x)] + \log[P_2(y|x, y > 0)] \right\}
\]

\[
= \left\{ \sum_{y=0} \log[1 - P_1(y > 0|x)] + \sum_{y>0} \log[P_1(y > 0|x)] \right\} + \left\{ \sum_{y>0} \log[P_2(y|x, y > 0)] \right\}
= \log L_1 + \log L_2
\]  

(14)

The two parts are therefore assumed to be a result of two different processes. The economic interpretation of the hurdle model is that absence is seen as a two-stage decision process where the first decision governs the decision to report sick and the second decision the length of the spell. In the present application the first process thus governs the decision to report sick for more than 14 days, and the second process the length of the spell.

Two models are needed in order to estimate the hurdle model. The first part of the model is usually specified as a logistic regression for the dichotomous outcome of being absent or attending. In the second part an appropriate count data model needs to be estimated.
5.2 A panel data approach

This analysis is based on a longitudinal, or panel, data set. A panel data set is a combined cross section and time series data set, i.e., individuals are followed over several years. There are several advantages of using panel data. The first one is that the number of observations increases which gives more variability and more efficiency. The second one is that the panel structure makes it possible to control for unobserved effects, which arise due to omitted variables. Modelling differences in behaviour across individuals is therefore considerably more flexible compared to cross section data (Greene, 2003).

Consider a Bernoulli trial where on a given day, an individual is absent with probability $p$ and at work with probability $(1-p)$. This trial is repeated $n$ times in a given year. Assuming these trials are independent, days absent follow a Poisson distribution with parameter $\lambda_i$ which is related to the covariates $x_i$. The Poisson distribution is thus given by

$$P(Y_i = y_i|x_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad y_i = 0,1,2,...$$

(15)

The most commonly formulation for $\lambda_i$ is in loglinear form such that $\lambda_i = \exp(x_i' \beta)$. The conditional mean for the Poisson panel data model is given by;

$$\lambda_{it} = \exp(x_{it} \beta + u_i)$$

(16)

where $i = 1,...,n$ indexes individuals and $t = 1,...,T$ indexes time, $x$ is a vector of covariates, $\beta$ is a vector of coefficients to be estimated and $u_i$ is an individual specific error term. The Poisson model is a non-linear model, and basic assumptions of the model is somewhat different than from an ordinary linear model. It can be shown that the conditional variance will equal the conditional mean. Furthermore, events are assumed to occur independently.

The individual specific error term in (16) represents unobserved individual effects. In order to control for these unobserved effects one needs to decide on whether to use fixed or random effects. The fixed effects estimation is used when the omitted variables (included in the individual specific error term) are correlated with the covariates. This is also known as unobserved heterogeneity. The random effects estimator is appealing when the unobserved effect is assumed to be uncorrelated with the covariates.
There are several reasons as to why a Poisson model might be too restrictive in this setting. Firstly, the conditional mean of $Y$ must equal the conditional variance. Failing to account for this may lead to too small standard errors of the coefficients (Cameron & Trivedi, 1986). If unobserved heterogeneity exists a phenomenon called overdispersion is likely to arise (Winkelmann, 1997). This yields a conditional variance that exceeds the conditional mean. A second reason to why a Poisson model might be inappropriate is the assumption of independence of spells. It is intuitive to assume that past absence is correlated with future absence. A more general count data model relaxing these assumptions is the negative binomial model. Hausman et. al. (1984) provides the panel data negative binominal model. If we let $\exp(\mu_i)$ in the Poisson model be gamma distributed with mean 1.0 and variance $1/u$, the panel negative binominal model results

$$P(y_{it}, \ldots, y_{it}) = \frac{\Gamma(\sum_{i} \lambda_{it}) \Gamma(\sum_{i} y_{it} + 1)}{\Gamma(\sum_{i} \lambda_{it} + \sum_{i} y_{it})} \times \prod_{t} \frac{\Gamma(\lambda_{it} + y_{it})}{\Gamma(\lambda_{it}) \Gamma(y_{it} + 1)}$$

(17)

where $\Gamma$ is the gamma function. In the negative binomial random effects estimation, the individual specific error term need not be uncorrelated with the covariates. If the uncorrelatedness of the covariates and the unobserved effects can be maintained, the random effects model is attractive (Greene, 2003). The negative binomial model is log-linear, which means that a coefficient needs to be exponentiated in order to obtain the estimated multiplier effect a one-unit change in a covariate has on the expected number of sick days.

In the first part of the hurdle model a panel data logistic regression is used. See appendix 1 for details of the logistic panel data regression. Ideally, in the hurdle model, the truncated at zero negative binomial (or Poisson) would have been used. However, the panel data version of this model does not exist in Stata 8.0 which is used in this study. The negative binomial model in (17) is therefore used. This is similar to the approach in Brown et. al. (2006) who also did not use a zero truncated model for the second part of the hurdle model.
6. Data

6.1 Study population

This study is based on data from the Diabetes Incidence Study in Sweden (DISS) and Statistics Sweden. The DISS study has since 1983 registered the incidence and the development of diabetes related complications, in individuals with onset of type 1 and type 2 diabetes at 15 to 34 years of age. Approximately 400 individuals are diagnosed each year of which approximately 75 percent with type 1 diabetes (Jonsson et al, 2000). About 65 percent of the registered type 1 diabetics are men, and 35 percent women (Blohmé et al, 1992). The level of ascertainment is estimated to be 90 percent for type 1 diabetics (Littorin et al., 1996).

For each individual in the DISS study four control persons matched on age, sex and county of residence are also included. In order to obtain information on sickness absence, the data in the DISS study is matched with the LISA data base compiled by Statistics Sweden. The complete data set contains information on the number of days for which the individual receives sickness insurance as well as other socioeconomic and demographic characteristics. In the present study, individuals diagnosed with type 1 diabetes and a control group are followed between 1993 and 2005. The year of diagnosis could be any between 1983 and 2005. The individuals were born between 1948 and 1970 and mean age at diabetes onset is at the age of 27.

The total sample consists of 17,998 individuals in 1993\(^{10}\). Several individuals are however excluded from the analysis. Firstly, as the analysis focuses on sickness absence from work only those employed are included. Anyone unemployed for more than 30 days in a given year is therefore excluded. Students, self-employed, individuals on maternity leave and individuals on disability pension are also excluded. The reason for these exclusions is to limit heterogeneity that might arise due to differences in the social security system. The social security system rules are different for self-employed and students. Individuals on maternity leave are entitled to sick pay but are excluded due to the fact that they are not working per se while on maternity leave. Individuals on disability pension are not entitled to sick pay.

After the exclusion 8,802 individuals remain in the sample in 1993. In 1993 there were 7,410 control persons and 1,392 diabetics while 6,731 control persons and 1,599 diabetics in 2005.

\(^{10}\) This number varies somewhat over the years a result of e.g. individuals moving abroad or deceasing.
6.2 The dependent variables

Since a hurdle model is employed in this thesis, two dependent variables will be specified. The first one is the number of days for which an individual received sickness insurance in a given year. For brevity, the number of days for which an individual receives sickness insurance will be referred to as days absent. Days absent are measured as calendar days, not working days. In Sweden, the employer is responsible for reimbursement during the first 14 days of a sickness spell. The data used in this analysis contains no information on sick pay from the employer which limits the data to include only absence spells lasting longer than 14 days. Table 2 shows descriptive statistics for the number of days absent.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>sickdays</td>
<td>100988</td>
<td>14.3</td>
<td>60.1</td>
</tr>
</tbody>
</table>

Table 2 Descriptive statistics for the dependent variable

Figure 4 shows average days absent for diabetics and the control group. Clearly, average days absent is higher for diabetics than for the control group in each year. Average days absent increases considerably for both groups during the 21st century, compare with figure 1.

![Figure 4: Average days absent](image)

The dependent variable in the logistic regression is whether the individual is absent or not. The great majority of the individuals are not absent. Figure 5 shows the proportion absent in each year by diabetics and the control group. More than 80 percent of the individuals are not...
absent in each year. The proportion absent in each year is substantially higher for the diabetics than for the non diabetics.

![Figure 5 Proportion absent](image)

### 6.3 The independent variables

The independent variables are based on the theoretical models and selected after data availability. Table 3 shows a description of the variables included in the analysis and table 4 descriptive statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>The age of the individual</td>
</tr>
<tr>
<td>woman</td>
<td>Dummy variable 0=man, 1=woman</td>
</tr>
<tr>
<td>income</td>
<td>Annual income minus average yearly tax rate. In 1000 SEK</td>
</tr>
<tr>
<td>edup</td>
<td>Dummy variable 1= highest educational level is upper secondary school</td>
</tr>
<tr>
<td>eduniv</td>
<td>Dummy variable 1= highest educational level is university</td>
</tr>
<tr>
<td>unemp</td>
<td>Average national unemployment rate</td>
</tr>
<tr>
<td>sickpay_level</td>
<td>Reimbursement rate from table 1</td>
</tr>
<tr>
<td>diab</td>
<td>Dummy variable 0=no diabetes in current year, 1=diabetes in current year</td>
</tr>
<tr>
<td>time_diag</td>
<td>Number of years since diagnosis</td>
</tr>
<tr>
<td>past_absence</td>
<td>Dummy variable 0=not absent the previous year, 1=absent the previous year</td>
</tr>
</tbody>
</table>

Note: omitted category of the education dummies is “highest educational level is elementary school”

**Table 3 Description of independent variables**
Table 4 Descriptive statistics for independent variables

The data does not contain any information on neither contracted working hours nor actual hours worked. It is therefore not possible to calculate the wage rate. As there is no proxy for working hours, annual net income will be used instead. Annual net income is calculated by subtracting the average national tax rate annually. The coefficient of annual income is expected to be negative. This is due to an inverse relationship between the wage and days absent. Under an efficiency wage approach the employer sets a wage above the market clearing wage as a method of controlling shirking (Barmby et. al., 1994). The intuition is that a higher wage induces not to shirk by raising the costs of being absent. The efficiency wage stands in proportion to the level of monitoring. If monitoring devices are readily available, the employer can set a lower wage. Moreover, in the reservation level of sickness, a higher wage increases the reservation level of sickness which would decrease expected absence.

The same argumentation could be extended to the relationship between unemployment and sickness absence. The unemployment rate can thus also be seen as a monitoring device as the cost of being absent increases in times of high unemployment. The increased cost of absence is due to the fact that it would be more difficult finding a new job if being dismissed in times of high unemployment (Johansson & Palme, 1996). Therefore, sickness absence should decrease in times of high unemployment. In order to control for this factor the average annual unemployment rate is included.

It is a well known fact that sickness absence differs between women and men. As women tend to have more days absent than men, the sign on this variable should be positive. Age is also included in the analysis. According to the health capital model, the depreciation rate

<table>
<thead>
<tr>
<th>Variable</th>
<th>1993 Diabetics (mean/percent)</th>
<th>Control group (mean/percent)</th>
<th>2005 Diabetics (mean/percent)</th>
<th>Control group (mean/percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>31</td>
<td>30</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>woman</td>
<td>0.31</td>
<td>0.30</td>
<td>0.31</td>
<td>0.32</td>
</tr>
<tr>
<td>income</td>
<td>113 600</td>
<td>113 800</td>
<td>196 700</td>
<td>201 400</td>
</tr>
<tr>
<td>edup</td>
<td>0.56</td>
<td>0.57</td>
<td>0.53</td>
<td>0.54</td>
</tr>
<tr>
<td>eduniv</td>
<td>0.26</td>
<td>0.26</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>past_absence</td>
<td>n/a</td>
<td>n/a</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td>time_diag</td>
<td>7</td>
<td>n/a</td>
<td>17</td>
<td>n/a</td>
</tr>
<tr>
<td>unemp</td>
<td>0.082</td>
<td>0.082</td>
<td>0.058</td>
<td>0.058</td>
</tr>
</tbody>
</table>

Note: n/a – not applicable
increases with age which causes health to deteriorate (Grossman, 1972). As less health capital leads to more sick days, age would be positive.

As can be seen in table 1, the reimbursement rate changed several times during the study period. In order to account for these changes, a variable indicating the actual reimbursement rate in the current year is included. In the reservation level of sickness, a higher reimbursement rate will increase expected absence.

The health stock is controlled for by including a dummy variable indicating whether the individual is a type 1 diabetic or not. This dummy variable takes on the value 1 at diagnosis which means that all individuals are treated as control persons before diagnosis. To furthermore control for the effect of diabetes, time since diagnosis is included as well. In figure 6 average sick days is plotted against time since diagnosis for selected years. It is rather difficult to determine the exact relationship between sick days and time from diagnosis.

![Figure 6](image)

**Figure 6** Average sick days for the number of years since diagnosis

In order to test whether previous absence spells is a predictor of new absence spells, a dummy variable indicating if the individual was absent in the previous year is also included. It is hypothesised that previous absence increases absence in the present year. Consequently, the sign on this variable should be positive.
7. Results

7.1 Model specification

The approach taken here is to first estimate the probability of being absent with a panel data logistic model. With panel data, it is possible to control for all unobserved, time-constant factors that affects sickness absence. As explained in chapter 5 this is called unobserved heterogeneity. It is reasonable to assume unobserved heterogeneity in the data which calls for the use of fixed effects. Unfortunately, it is not possible to use a Hausman test in order to test whether to use fixed or random effects estimation (Greene, 2003). A fixed effects model is therefore assumed a priori. A drawback with the fixed effects estimation is that no time invariant variables can be used, such as sex, since the time invariant variables are subtracted away in the estimation procedure. Yet another drawback is that the estimation algorithm excludes those individuals for which the dependent variable never changes. This means that those individuals absent in each year, or never absent over the years, are not included in the estimation sample. The random effects estimation does not suffer from these drawbacks but as a fixed effects model seems more reasonable, a random effects estimation was not used. The model fit of the fixed effects model is assessed by the percent correctly predicted\(^{12}\). The model predicts 30 percent rightly, which is only moderate. However, this goodness of fit test is only fairly reliable (Wooldrige, 2002). A pseudo R\(^2\) is also calculated which gives a value of 11 percent which confirms the moderate fit of the model. It should however be pointed out that poor predictability is a common problem in social sciences.

For the second part, a natural starting point would be a Poisson model. There are several indicators of this model being inappropriate. Firstly, as table 2 indicates, the variance exceeds the mean substantially. Secondly, the assumption of occurrence independence is clearly inappropriate when analysing sickness absence. To account for these aspects, a negative binomial model is instead estimated. A Hausman test\(^{13}\) is performed in order to test the hypothesis of independence between the random effects and the covariates. A rejection of the

\(^{12}\) In this test, predicted probabilities are compared with the actual outcome. If the predicted probability of a positive outcome is larger than 0.5 the individual is ascertained a 1. If the predicted probability of a positive outcome is less than 0.5 the individual is ascertained a 0. The percentage of times the predicted \(y_i\) matches the actual \(y_i\) is the percent correctly predicted.

\(^{13}\) The Hausman test; \(H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' \text{var}(\hat{\beta}_{FE}) - \text{var}(\hat{\beta}_{RE})^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE})\) where \(\hat{\beta}_{FE}\) is a vector of fixed effects estimates and \(\hat{\beta}_{RE}\) is a vector of random effects estimates. \(H\) is \(\chi^2\)-distributed with degrees of freedom equal to the number of time-variant variables.
hypothesis of independence indicates that a fixed effects model should be estimated. The Hausman test rejects the hypothesis of random effects and a fixed effects model is estimated. Allison and Waterman (2002) show that the fixed effects negative binomial model is not in fact a “true” fixed effects model, i.e., it does not control for all time-invariant covariates. It is therefore possible to include time invariant variables in the fixed effects negative binomial model. Hence, sex is included in this model. A drawback with this model is, as in the logistic model, the fact that when an individual does not change state, that individual does not contribute to the likelihood function and is therefore excluded. This means that individuals absent each year are excluded from the estimated model. The fit of the negative binomial model is assessed with a graph where the probabilities from the negative binomial model are plotted against the observed proportions. Figure 7 shows that the negative binomial model substantially over predicts the lowest values. Moreover, a large proportion of the study population receives sickness insurance for the whole year which the fitted negative binomial model does not account for.

Standard errors from the *xtlogit* and the *xtnbreg* commands in *Stata 8.0* cannot be adjusted for heteroskedasticity as there is no option for robust standard errors, nor can autocorrelation be accounted for. One solution to this is of course to ignore the panel data and use ordinary cross sectional methods. This is thus a trade-off between correct standard errors and the advantages of using panel data. By neglecting heteroskedasticity and autocorrelation standard errors will be incorrect which could result in incorrect hypothesis tests and confidence intervals. The advantage of using panel data is more variability, degrees of freedom and efficiency. In addition, panel data makes it possible to control for individual heterogeneity which, if neglected, could give biased estimates (Baltagi, 2005).
7.2 Estimation results

The estimated coefficients from the two estimated models in the hurdle model are shown in table 5. Marginal effects for the negative binomial model are shown in table 6, but could not be estimated for the logistic model. A correlation matrix for the included variables is found in appendix 2.

7.2.1 The logit model

The coefficients in the first part of the hurdle model are overall significant, and have the sign expected from theories. The income variable suggests that a rise in income decreases the odds of being absent. As it is somewhat difficult to comprehend the exact meaning of an odds that decreases, it would be easier to interpret this odds as the odds of not being absent which is just the inverse\textsuperscript{14}. The odds of not being absent thus increases with 1.16 for a unit increase in income. This confirms that a higher wage implies a higher reservation level of sickness and less expected absence. In addition, the efficiency wage hypothesis, that a higher wage encourage not to absent, is thus confirmed. Moreover, a higher unemployment rate also decreases the odds of being absent as expected. As predicted from the labour-leisure model, a higher sick pay increases the odds of being absent.

Age enters significantly and indicates that the odds of being absent increases with age. This result is consistent with Grossman’s (1972) suggestion on that the depreciation rate increases with age. A higher depreciation rates yields a lower optimal health stock and consequently more days lost to illness. Age is thus also a measure of health status. Grossman also suggests that higher education yields a higher optimal health stock and fewer days lost to illness. Although insignificant, the odds ratios show that a higher education decreases the odds of being absent.

Being absent the previous year increases the odds of being absent in the present year 1.5 times. The estimation procedure ignores all individuals who never change state, which means that those absent all years, or never absent, are not included. This could overstate the effect of this variable for which reason the absolute effect of this variable should be carefully interpreted. The odds ratio for the woman variable could unfortunately not be estimated due to

\textsuperscript{14} \frac{1}{0.86} = 1.16
<table>
<thead>
<tr>
<th>Variable</th>
<th>Log model</th>
<th>Neg bin model</th>
</tr>
</thead>
<tbody>
<tr>
<td>woman</td>
<td>-0.14***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td></td>
</tr>
<tr>
<td>income</td>
<td>0.86***</td>
<td>-0.1***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>age</td>
<td>1.09***</td>
<td>0.022***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>edup</td>
<td>1.21</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>eduniv</td>
<td>0.79</td>
<td>0.28***</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>unemp</td>
<td>5.75^-7***</td>
<td>-10.59***</td>
</tr>
<tr>
<td></td>
<td>(9.6*^-7)</td>
<td>(0.79)</td>
</tr>
<tr>
<td>sickpaylevel</td>
<td>691***</td>
<td>-2.4***</td>
</tr>
<tr>
<td></td>
<td>(624.1)</td>
<td>(0.54)</td>
</tr>
<tr>
<td>past_absence</td>
<td>1.46***</td>
<td>0.24***</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>diab</td>
<td>6.9***</td>
<td>-0.18**</td>
</tr>
<tr>
<td></td>
<td>(1.42)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>time_diag</td>
<td>0.77***</td>
<td>0.0089*</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>time_diagsq</td>
<td>1.01***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>diab_income</td>
<td>0.972***</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>constant</td>
<td>2.9***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observations</th>
<th>37 876</th>
<th>8854</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals</td>
<td>4997</td>
<td>2867</td>
</tr>
<tr>
<td>Loglikelihood</td>
<td>-11910</td>
<td>-30 358</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses, *** p<0.001, ** p<0.05, * p<0.1

*Table 5 Estimated hurdle model coefficients*
Table 6 Negative binomial marginal effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Marginal effects for the neg bin model</th>
</tr>
</thead>
<tbody>
<tr>
<td>woman</td>
<td>-0.24</td>
</tr>
<tr>
<td>income</td>
<td>-0.15</td>
</tr>
<tr>
<td>age</td>
<td>0.038</td>
</tr>
<tr>
<td>edup</td>
<td>-0.008</td>
</tr>
<tr>
<td>eduniv</td>
<td>0.53</td>
</tr>
<tr>
<td>unemp</td>
<td>-18</td>
</tr>
<tr>
<td>sickpaylevel</td>
<td>-4.1</td>
</tr>
<tr>
<td>diab</td>
<td>-0.3</td>
</tr>
<tr>
<td>diab_income</td>
<td>-0.0004</td>
</tr>
<tr>
<td>time_diag</td>
<td>0.015</td>
</tr>
<tr>
<td>past_absence</td>
<td>0.41</td>
</tr>
</tbody>
</table>

above stated reasons. It should however be emphasized that the effect of being a woman is included in the estimation of the fixed effects.

As with the income variable, the interaction between diabetes and income show that the odds of being absent decreases. Again, this would be easier to interpret with the odds of not being absent\(^{15}\). The odds of not being absent increases with 1.03 for a one unit increase in income for diabetics. A rise in income for diabetics increases the odds of not being absent less than a rise in income for the whole sample. This suggests that diabetics are not as sensitive to changes in income. A possible explanation for this could be that income is less important if sick. Furthermore, as can be seen in table 4, mean income for diabetics is lower than for the control group. This implies a lower reservation level of sickness and thus higher expected absence. In an efficiency wage approach this would create fewer incentives to attend work. As explained in the theoretical discussion, the wage trend for diabetics could be expected to be worse than for non-diabetics. Income can thus also be seen as a proxy for health. The lower income and the lower sensitivity for income changes confirm the hypothesis of higher probability of sickness absence.

The diabetes variable and the time since diagnosis variable capture the effects of health status on sickness absence. The diabetes variable is highly significant and shows that being a diabetic increases the odds of being absent almost sevenfold. This is in line with the predictions from the labour-leisure model. Time since diagnosis and its quadrate are also included and are both highly significant. Time since diagnosis has a negative effect on the odds of being absent and its quadrate has a positive effect, which indicates a parabolic relationship in a life time perspective. This means that with only a few years from diagnosis

\[^{15}\] 1/0.972 = 1.03
the odds of being absent decreases but the more years since diagnosis the higher the odds of being absent.

7.2.2 The negative binomial model

The coefficients in the negative binomial model are overall significant. For ease of interpretation, table 6 gives the estimated marginal effects for the negative binomial model. As previously explained, in the fixed effects negative binomial model, it is possible to estimate the effect of sex, a time invariant variable. The coefficient is highly significant and negative suggesting that men are more absent than women. The marginal effect is rather small, still women have fewer days absent than men. This was not expected.

The prediction from the labour-leisure model, that a higher income decreases expected absence, is confirmed in this part of the model as income has a negative effect on sickness absence. The unemployment variable is, as expected, negative and indicates fewer days absent in times of high unemployment. As table 6 suggests, the unemployment rate contributes with the largest marginal effect on the number of days absent. The sick pay level however does not have the expected sign. The variable indicates that for individuals with days absent these would decrease with an increase in the reimbursement rate. This could be a result of a substantially less percentage absent during the beginning of the 90’s when the reimbursement rate was high.

Age enters significantly and confirms that as health deteriorates the number of days absent increases. The education variables show a rather peculiar effect. The upper secondary school dummy is negative while the university dummy is positive suggesting that a higher education increases expected absence. The upper secondary school dummy is however insignificant while the university variable is significant. A possible explanation for this could be that with higher education, salary will also be higher. The income effect suggests that high income earners can afford to report sick. Past absence is also significant, suggesting that being absent the previous year increases days absent the present year.

The diabetes variable is significant and negative. This would be interpreted as, conditional on being absent, the fact that you are diagnosed with type 1 diabetes decreases the number of days absent. This is not what would be expected from the health capital model. A possible explanation for this is that the control persons with long absence spells also are sick. In this
study only absence spells lasting longer than 14 days are analysed. A doctor’s certificate is needed if absent more than six consecutive days, which means that all individuals are being absent for medical reasons\textsuperscript{16}. Thus, the health status of the individuals in the control group with long absence spells is not known but could be assumed to be lower than the general population. Thus, compared to the control group, who also suffer from illness, diabetes per se does not increase length of the spell. It should however be pointed out that the marginal effect of being a diabetic is small.

A graphical examination shows a linear relationship between years since diagnosis and days absent. The square of years since diagnosis is therefore not included. The variable years since diagnosis is significant and positive, indicating an increase in days absent with the number of years since diagnosis. This effect is somewhat peculiar since the diabetes variable was negative. It suggests, however, that diabetes in fact does have some influence on the number of days absent.

The interaction between diabetes and income is insignificant and negative. As in the logistic model, the effect of diabetes income on sickness absence is smaller than the effect of income for the entire sample. In line with the above reasoning, income seems to affect diabetics less than when considering the entire sample.

7.3 Sensitivity analysis and robustness checks

In this study, only those unemployed less than 30 days are included. This restriction is checked by altering this figure. Fewer or more days does not affect the significance of variables and produces little changes in the magnitude of the variables. The only exception is in the negative binomial model where the variable time since diagnosis becomes insignificant when those unemployed more than 30 days are included.

The data set consists of four control persons for every diabetic. After the exclusion of individuals, this might not be true anymore. Ideally, one would exclude the matching control persons if the diabetic is excluded. In order to check the consequence of the exclusion rules, the proportion of diabetics and control persons are controlled before and after the exclusion. The result shows that the percentage of diabetics is the same after the exclusion as before the exclusion.

\textsuperscript{16} This can most certainly be questioned. It could be that the individual influence the physician to certificate absence. That is, shirking could be possible also in the case of long absence spells.
exclusion. It could also be the case that a specific exclusion rule has higher proportion diabetics. This is the case for individuals in the category disability pension, which consists of a higher proportion diabetics. Even though these individuals are not on sick leave, it underestimates the impact of diabetes on absence from the work force.

In the fixed effects estimation of the logistic and the negative binomial model, individuals who never change states are excluded from the analysis. Individuals absent each year, or never absent, are thus excluded from the analysis. This might impact the result of the fixed effects estimation, if e.g. a higher proportion diabetics than control persons are absent each year. The random effects estimation of the logistic and the negative binomial models are therefore given in appendix 3. The result of the random effects estimation differs somewhat from the fixed effects estimation. Most importantly, the diabetes variable becomes insignificant in the negative binomial model. It should be remembered that the Hausman test rejects the negative binomial random effects model. It is difficult to determine if the changes in coefficients is due to the random effects estimation or the inclusion of more individuals. Most likely, both effects affect the estimation.
8. Discussion and conclusion

8.1 Discussion of the results

The aim of this study was twofold. Firstly, a theoretical explanation of sickness absence after the onset of type 1 diabetes was given. The main conclusion from the theoretical discussion was that the onset of type 1 diabetes will alter preferences and behaviour. Diabetes was due to an increased mortality rate assumed to lower optimal health stock which would lead to more sick days in a given year in the health capital model. The level of the sickness index would increase permanently in the labour-leisure model, increasing marginal utility of leisure and the probability of exceeding the reservation level of sickness. There are however some aspects of the theoretical models which could be questioned. The health capital model assumes all activities to be mutually exclusive. This is not particularly realistic, especially since a diabetic most certainly has to e.g. maintain health with insulin injections during work time. Moreover, sick time being regarded as completely lost time is also unrealistic. In the labour-leisure model, these assumptions are relaxed. As also concluded in Brown and Sessions (1996), the strength of the labour-leisure model is that health is continuous and not dichotomous, where healthy individuals would work and sick would absent. The advantage of this approach is that not only healthy individuals face the decision between work and absence. The complexity of absence decisions is thus well captured in the labour-leisure model. This is the main contribution of the labour-leisure model in this study. The labour-leisure model demonstrates that sickness absence is not just about a lower optimal health stock but also influenced by other factors such as sick pay, work hours and the wage rate through the reservation level of sickness.

Secondly, an empirical analysis was conducted in order to assess whether sickness absence differed between type 1 diabetics and a control group. As predicted from the theoretical discussion, the empirical analysis demonstrated that diabetics had a sevenfold excess risk of sickness absence, ceteris paribus. Moreover, the theoretical discussion predicted longer absence spells for diabetics. This was not confirmed in the empirical analysis. As previously concluded, this is assumed to be a result of the control group also being stricken with illness.

Time since diagnosis seems to be an important explanation for sickness absence in diabetics. In the logistic model, time since diagnosis is modelled as a second-order polynomial, even
though it is rather difficult to discern the exact relationship. There seems to be a peak in absent days around ten years from diagnosis, which would be when diabetes related complications often appear. In both the logistic and the negative binomial model, it is confirmed that absence increases with time from diagnosis. This suggests that in a life time perspective, health will deteriorate even further. Moreover, diabetics seem to be less sensitive with regard to the effect of income on absence. It should also be pointed out that there is a spurious relationship between income and days absent. According to the labour-leisure model, income affects days absent but obviously, income is also affected by absence from work. This could be true especially for diabetics if higher absenteeism leads to worse career opportunities for the diabetic.

This study clearly indicates that the onset of type 1 diabetes implies increased sickness absence. An effect on the length of the spell could however not be predicted. The sensitivity analysis revealed a higher proportion diabetics being on disability pension. As it is not possible to be on sick leave while on disability pension, these individuals were excluded. However, disability pension is in fact preceded by sick leave. The effect of diabetes on absence from work is therefore underestimated.

The result of other studies examining sickness absence varies. It is however difficult to compare the results of this study with other studies, as there is a vast difference in study groups and methodologies applied. Moreover, absence behaviour is modelled with Swedish data. Several studies also suggest that prevailing macroeconomic circumstances and insurance systems affects sickness absence. Hence, it might not be possible to generalise the results of this study to settings too different.

The strength of this study is the rich and reliable data upon which it is based. Few measurement errors could be expected in register data compared to self-reported data. Moreover, the DISS study covers 90 percent of the type 1 diabetics diagnosed between 1983 and 2005 in Sweden which gives reliable estimates. The longitudinal data also gives more reliable estimates as the estimation takes account of individual changes in absence spells over time.

8.2 Study limitations

Due to lack of data on working hours, it was not possible to calculate the wage rate. Instead, annual income was used as a measure of the cost of being absent. This would impact the
result only if there are individuals not working full time, which is rather likely. For those working part time the cost of being absent will be overestimated. The result will be biased only if diabetics are overrepresented as part time workers. Changes in the unemployment rate and the reimbursement rate from sickness insurance are controlled for but not other macroeconomic fluctuations and changes in sickness insurance. If these changes affects diabetics and non diabetics the same this would not matter. The impact of supplementary insurance is also not controlled for.

8.3 Policy implications and suggestions for future research

This study provides evidence of increased sickness absence in type 1 diabetics. This affects not only the diabetic but also society, in terms of costs for social insurance and lost production. It is therefore to the benefit of all to reduce sickness absence in diabetics.

The present study does not investigate sickness absence before the onset of type 1 diabetes. It would be interesting to examine when health starts to deteriorate and if this leads to increased sickness absenteeism. It would also be interesting to extend the analysis by including the effects of diabetes severity and diabetes control. This could be done by e.g. examining the effects of presence of complications and blood sugar control.

8.4 Conclusion

This study offers a theoretical explanation of sickness absence behaviour in the presence of a chronic illness such as type 1 diabetes. The main conclusion from the theoretical discussion is that the onset of type 1 diabetes will alter preferences and sickness absence behaviour. Diabetes was due to an increased mortality rate assumed to lower optimal health stock which resulted in more sick days in a given year in the health capital model. The level of the sickness index increased permanently in the labour-leisure model, increasing marginal utility of leisure and the probability of exceeding the reservation level of sickness. The empirical analysis show that type 1 diabetics are exposed to a sevenfold excess risk, ceteris paribus, of being absent compared to the control group. However, being stricken with diabetes does not increase the length of the spell. In a life time perspective sickness absence will increase with time from diagnosis. In conclusion, the onset of type 1 diabetes implies increased sickness absence.
Literature


Borch-Johnsen K., Kreiner S., Deckert T. (1986). ‘Mortality of Type 1 (insulin-dependent) diabetes mellitus in Denmark: a study of relative mortality in 2930 Danish Type 1 diabetic patients diagnosed from 1933 to 1972’. *Diabetologia* 29 767-772


National Institute of Economic Research
[http://www.konj.se/download/18.988707fc16e9df3b7ff6892/KLj03_ruta3.pdf](http://www.konj.se/download/18.988707fc16e9df3b7ff6892/KLj03_ruta3.pdf) 2008-04-03


The Swedish Diabetes Association [www.diabetes.se](http://www.diabetes.se) 2007-12-11

The Swedish Social Insurance Agency

Appendix 1: The panel data logistic model

The panel data logit model (Greene, 2003) is represented by the following equation

\[ y_{it} = x'_{it} \beta + v_{it} + u_i, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T \]

where \( y_{it} \) equal 0 or 1, \( x \) is a vector of covariates, \( \beta \) is a vector of coefficients, \( v_{it} \) is the residual and \( u_i \) is the unobserved individual error term. Fixed effects arise if \( u_i \) and \( x_{it} \) are correlated and random effects if uncorrelated. The fixed effects model is

\[ y_{it} = \alpha_i d_{it} + x'_{it} \beta + \varepsilon_{it} \]

where \( d_i \) is a dummy variable which takes the value one for individual \( i \) and zero otherwise. The parameters to be estimated are the \( K \) elements of \( \beta \) and the \( n \) individual terms. Hence, the number of parameters to be estimated could be huge. The problem with the fixed effects logistic model is that \( u_i \) cannot be consistently estimated, which is known as the incidental parameters problem. The parameters in the logistic model will then be biased.
Appendix 2: Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>sickdays</th>
<th>sex</th>
<th>income</th>
<th>age</th>
<th>edup</th>
<th>eduniv</th>
<th>unemp</th>
<th>sickpay</th>
<th>diab</th>
<th>Diab_income</th>
<th>time_diag</th>
<th>time_diagsq</th>
<th>past_absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>sickday</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sex</td>
<td>0.0919</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>income</td>
<td>-0.2214</td>
<td>-0.1829</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>0.0756</td>
<td>0.0460</td>
<td>0.2028</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edup</td>
<td>0.0428</td>
<td>0.0072</td>
<td>-0.1501</td>
<td>-0.0636</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eduniv</td>
<td>-0.0683</td>
<td>0.0233</td>
<td>0.2777</td>
<td>0.0184</td>
<td>-0.7143</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unemp</td>
<td>-0.0942</td>
<td>-0.0137</td>
<td>-0.2079</td>
<td>-0.4679</td>
<td>0.0081</td>
<td>-0.0246</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sickpay</td>
<td>0.0193</td>
<td>0.0076</td>
<td>0.0124</td>
<td>0.0200</td>
<td>0.0055</td>
<td>-0.0053</td>
<td>-0.3389</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>diab</td>
<td>0.0487</td>
<td>0.0145</td>
<td>-0.0127</td>
<td>0.0335</td>
<td>0.0013</td>
<td>-0.0137</td>
<td>-0.0333</td>
<td>0.0016</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>diab_income</td>
<td>-0.0484</td>
<td>-0.0435</td>
<td>0.2231</td>
<td>0.0740</td>
<td>-0.0453</td>
<td>0.0683</td>
<td>-0.0764</td>
<td>0.0062</td>
<td>0.8072</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time_diag</td>
<td>0.0587</td>
<td>0.0170</td>
<td>0.0203</td>
<td>0.1493</td>
<td>-0.0048</td>
<td>-0.0097</td>
<td>-0.1146</td>
<td>0.0074</td>
<td>0.8893</td>
<td>0.7650</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time_diagsq</td>
<td>0.0606</td>
<td>0.0167</td>
<td>0.0349</td>
<td>0.1950</td>
<td>-0.0063</td>
<td>-0.0076</td>
<td>-0.1450</td>
<td>0.0114</td>
<td>0.7560</td>
<td>0.6767</td>
<td>0.9623</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>past_absence</td>
<td>0.5017</td>
<td>0.0950</td>
<td>-0.1552</td>
<td>0.0740</td>
<td>0.0532</td>
<td>-0.0904</td>
<td>-0.0837</td>
<td>-0.0165</td>
<td>0.0596</td>
<td>-0.0133</td>
<td>0.0611</td>
<td>0.0596</td>
<td>1.000</td>
</tr>
</tbody>
</table>
### Appendix 3: Random effects coefficients

<table>
<thead>
<tr>
<th>Variable</th>
<th>Logit model</th>
<th>Neg bin model</th>
</tr>
</thead>
<tbody>
<tr>
<td>woman</td>
<td>-0.06***</td>
<td>-0.08***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>income</td>
<td>0.90***</td>
<td>-0.08***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>age</td>
<td>1.03***</td>
<td>0.02***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>edup</td>
<td>1.14***</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>eduniv</td>
<td>0.0003***</td>
<td>0.23***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>unemp</td>
<td>0.00003***</td>
<td>-11.3***</td>
</tr>
<tr>
<td></td>
<td>(0.00003)</td>
<td>(0.57)</td>
</tr>
<tr>
<td>sickpaylevel</td>
<td>8841***</td>
<td>-2.9***</td>
</tr>
<tr>
<td></td>
<td>(6928)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>past_absence</td>
<td>5.98***</td>
<td>0.44***</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>diab</td>
<td>5.79***</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.69)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>time_diag</td>
<td>0.82***</td>
<td>0.0067***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>time_diagsq</td>
<td>1.01***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
<td></td>
</tr>
<tr>
<td>diab_income</td>
<td>0.97***</td>
<td>-0.01***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>constant</td>
<td>3.3***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td></td>
</tr>
</tbody>
</table>

Observations: 100983, 11853  
Individuals: 16172, 5866  
Loglikelihood: -27756, -64549