PREDICTING SICK LEAVE AND LENGTH OF LEAVE AMONG WORKERS: A HURDLE MODEL APPROACH

Economic theory

by

Sera Chiuchiarelli

(3160117)

Major Paper presented to the Department of Economics
of the University of Ottawa
to obtain a Master's degree

Supervisor: Professor Louis-Philippe Morin

ECO 7997

Ottawa, Ontario
May 2010

Abstract

This paper uses data from the Canadian Community Health Survey from 2003 and 2005 to investigate the determinants of sick leave. A two-step hurdle model is specified, involving a probit regression in part one and a zero-truncated negative binomial model in part two. By employing such an approach I am able to determine both the probability of taking a sick day, and the number of sick days an individual takes conditional on at least one absence. The findings suggest that regardless of gender, as individuals' age they are less likely to take a sick day but more like to take a longer leave. Individuals who earn higher incomes and have completed more schooling are less likely to take a sick day and they take less time off during an absence. Parents behave in the same way regarding sick leave as all other individuals. Finally, the results of my sensitivity analyses suggest that the model is both sensitive to variable specification and maximum number of sick days.

INTRODUCTION

According to Statistics Canada between 2004 and 2007, full-time workers in Canada lost an average of 7.8 days of work due to illness or disability. These types of absences are commonly referred to as sick leave or absenteeism. Alternately, there are times when we simply do not feel that it is necessary to take time away from the job, despite our condition and make our way into the office, regardless. This concept is commonly referred to as presenteeism. While sick leave is a relatively straightforward notion, the factors that affect one's decision to present themselves for work or remain at home are not always that simple.

In principle, an individual may base their decision solely on the way they feel physically or mentally, however, this may not always be the case. Some individuals may take into consideration their financial needs or the needs of their family. Some incentives that motivate presenteeism are unpaid sick leave and/or lack of health benefits or health insurance. Other motivations may include risk of unemployment or inability to take time off sick because of an impending deadline or meeting. On the other hand, some incentives for absenteeism may include lack of productivity, speedier recovery and reduced spread of illness or disease. In addition, an individual may choose to take a sick day based on another individual's health rather than their own.

Consequently, once the decision to take sick leave is made, a related issue is duration. While the two decisions may involve the same or similar factors, both decisions are not considered to be simultaneous, but rather sequential. Very few people wake up sick in the morning and decide to take one and only one day off sick. Typically, on the first day one decides whether or not to take a sick day and then remains home sick until they feel they are of adequate or complete health to resume their job without potentially infecting others. Some examples include bacterial and viral infections. According to the Public Health Agency of Canada, influenza can take anywhere between a week to ten days to recover from, providing there are no complications which can lead to more severe illnesses such as pneumonia (Public Health Agency of Canada, 2010). Similarly, the common cold is a viral infection affecting the respiratory system according to the Canadian Centre for Occupational Health and Safety can endure anywhere between two to seven days, though in some cases can last up to two weeks (Canadian Centre for Occupational Health and Safety, 2008).

The purpose of this paper is to determine the economic factors that contribute both to sick leave and to its length. Therefore, I define two-part or hurdle model using a probit model in the first part and a zero-truncated negative binomial model in the second part. This model will allow me to determine whether or not the characteristics of an individual who takes sick leave are equivalent to those that determine the length of their leave.

Unlike the existing literature which models sick leave based on children and other determinants, here I limit my variable selection in an attempt to reduce the amount of self-selection bias by excluding as many self-reported health or job related variables. Similarly, by performing a sensitivity analysis which excludes the dummy variable for chronic disease I am able to test the sensitivity of the model when health related variables (which may contribute to endogeneity bias) are eliminated from the model.

Overall my findings suggest that regardless of gender, as individuals age they are less likely to take a sick day but more like to take a longer leave. Individuals who earn higher incomes and have completed more schooling are less likely to take a sick day and they take less time off during an absence. I also find that the likelihood of taking a sick day and the length of a sick leave is the same for people with children under the age of 12. Lastly, the results of my sensitivity analyses suggest that the model is both sensitive to variable specification and maximum number of sick days.

The remainder of the paper is organized into nine sections. Section 2 provides a summary of some existing literature on the determinants of sick leave. Section 3 includes a brief background regarding the Canadian Community Health Survey and describes the data set I use. Section 4 describes the hurdle model and why it is useful. Section 5 presents my findings, while Section 6 discusses potential problems or limitations and provides some sensitivity analyses. Section 7 summarizes the preceding sections and provides some closing remarks.

LITERATURE REVIEW

Given that sick leave is the dependent variable in my model, I limit my analysis to only those articles which model sick leave. Furthermore, I am most interested in the impact of children, so I pay particular attention to papers that pertain to both sick leave and children.

Marshall (2006) considers various factors contributing to sick leave in Canada. The author considers three data sets: Absence from Work Survey, 1979-1992; Survey of Labour and Income Dynamics, 1993-2003; and Labour Force Survey, 1979-2003. She considers a linear regression model with the following factors: sex, age, marital status, education level, stress, overall health, physical or mental disability, unionized or non-unionized employment, health insurance plans, public sector or private sector employment, permanent or temporary employment and good or service sector employment. Her findings suggest that health factors such as age, stress, unionized jobs, health insurance plans, and occupational health have a positive effect on the length of sick leave and have resulted in extended sick days.

In addition, given that some of the data she uses dates back to 1979, she is able to do some comparative analysis and finds that between 1979 and 2003 the overall percentage of workplace absences has decreased. Similarly, this coincides with the slowly decreasing retirement age and together both have had a negative effect on the length of sick leave.

Lastly, she investigates the effect of workplace safety campaigns, which have succeeded in decreasing the amount of workplace related injuries. However, her findings show that work related injuries still represent the majority of sick leaves lasting 17 weeks or more.

The literature review done by Allebeck and Mastekaasa (2004) considers many of the same variables Marshall (2006) does and many more from a list of nearly one hundred publications. They divide their analysis into subsections pertaining to specific classes of variables such as socio-demographic indicators, working conditions and health related research topics. The purpose of their research is to determine causal links between indicators and sick leave; hence they compare results across articles which they deem to be of high scientific quality. In particular, their study considers family structure. Their results suggest that marital status does not contribute to sick leave and the effect of children in the home is contradictory and therefore inconclusive. They do however find that sickness absence tends to be higher amongst women versus men, amongst older versus younger workers and in rural versus urban communities, though there is no evidence of a clear causal relationship. In addition, they find that there is a negative correlation between sickness and socio-economic status, but once again there is little causal relationship. In terms of labour market conditions they find that there is little causal evidence to support a negative correlation between unemployment and sick leave. Lastly, they consider the impacts of psychological and physical work environments on sick leave. They find there some evidence to support that various psychological stress factors such an overwhelming workload, conflict with supervisors or co-workers, changes in working conditions and job insecurity increases the number of sick days. However, they find little evidence to support any causal link between physical working environment and sick leave, except perhaps in the case of ergonomic load.[1]

Similarly, van Poppel et al. (2002) consider a statistical analysis on the accuracy of self-reported sick leave. For their analysis, they consider only individuals working in the cargo department at Schiphol airport, who handle luggage and other cargo by hand (i.e. those who would be prone to take sick leave due to work induced back pain) over a one year period. They provide participants with a monthly questionnaire regarding sick leave for the first six months and then two follow up surveys at nine and twelve months. Finally, they compare the results of these eight surveys against company records and compute interclass correlation coefficients to test for reliability. Comparing survey results with company records, only 241 individuals remain in the sample and account for 320 episodes of sick leave, over a one year period. Grouping the sick leave into two lengths short leave ($\leq$ 7 days) or long leave ($>$ 7 days) they find that individuals recall bouts of short leave more accurately than long leave. The authors therefore, recommend a recall period somewhere between two and six months would reduce bias caused by self-reported sick leave. Furthermore, they find that the duration of sick leave due to back

1 The authors refer to ergonomic load as physical working environments which involve physically demanding work, uncomfortable working position or heavy lifting.

pain was longer than for all other illness, which adds support to the article by Allebeck and Mastekaasa (2004).

Vistnes (1997) argues that it is necessary to include health indicators for the respondents in the sample. She uses data from the household component of the 1987 National Medical Expenditure Study and specifies a hurdle model. This two-part decision model is used to determine the outcome of sick leave, by implementing a logit model in the first stage and a zero-truncated negative binomial model in the second stage. From the first stage she is able to predict the likelihood of taking sick leave and in the second to predict the number of days in a year the respondents takes off. Her analysis includes the same variables across both decisions and focuses primarily on differences in sick leave by gender. In addition to controlling for children and health status, she also considers other indicators of health status such as chronic conditions, smoking habits and obesity. Likewise, she also includes a number of work related indicators such as job tenure, type of occupation and industry, as well as, firm size. With respect to women and children her results indicate that children below the age of six increase a woman's sick leave and increase the length of sick leave once they miss a day. Conversely, she finds that men's absences are not influenced by the presence of children.

Overall the studies tend to focus on many of the same variables which I will be considering here. Primarily, I focus on age, gender, chronic condition, employment status, education, income, and most importantly the impacts of children on sick leave. Like, Vistnes (1997) I have chosen to specify a two-step or hurdle model except I choose to specify a probit model in part one instead of a logit model. I choose to specify a probit instead because I assume that the underlying distribution of the error terms is known and is normal. Similar to Vistnes (1997), Allebeck and Mastekaasa (2004), I will be running the model by gender to see if any differences exist. Lastly, by using a hurdle model to predict my results, part one determines the likelihood of taking a sick day, while part two determines the number of sick days an individual takes conditional on them having taken a sick day. Therefore, by conducting a similar analysis to Vistnes (1997) in addition to comparing my findings against hers, I am able to test whether or not the determinants of each part of the decision are the same.

DATA

I use the Canadian Community Health Survey (CCHS) by Statistics Canada. This representative survey is ideal because it includes socio-demographic, household, income and labour indicators, in addition to measures of sickness and health conditions. This cross-sectional population survey was first conducted in 2001 and includes data collected over a two-year cycle. The first year of a cycle ".1" consists of a general population health survey and the second year of a cycle ".2" consists of a provincial survey on a focused health related topic involving a smaller sample size (Statistics Canada, 2005a, 2006a). For the purpose of this paper, I construct a repeated cross-section data set combining the Public Use Micro Data Files (PUMF) from Cycle 2.1 and Cycle 3.1.[2] For Cycle 2.1, the data collection began in January of 2003 and ended in December of 2003 and included respondents from 126 health regions[3] (Statistics Canada, 2005a). Similarly, for Cycle 3.1, in 2005 the data collection took place over the same twelve months, except this time it included only 122 health regions (Statistics Canada, 2006a). In addition, the

2 Cycle 1.1 was not included in the panel because the chosen dependant variable did not exist until Cycle 2.1.

3 Health regions used for the CCHS are based on the Census Sub-Divisions used in the 2001 Census. They correspond to smaller geographical areas within each province, except for the territories, which each represent a health region (Statistics Canada, 2005a, 2006a).

cycles target individuals 12 years and older living in private residents, in any one of the ten provinces or three territories. The two cycles excluded only those individuals who were living in an institution, a remote region, on Crown Land, on an Indian Reserve or who were members of the Canadian Armed Forces (Statistics Canada, 2005a, 2006a).

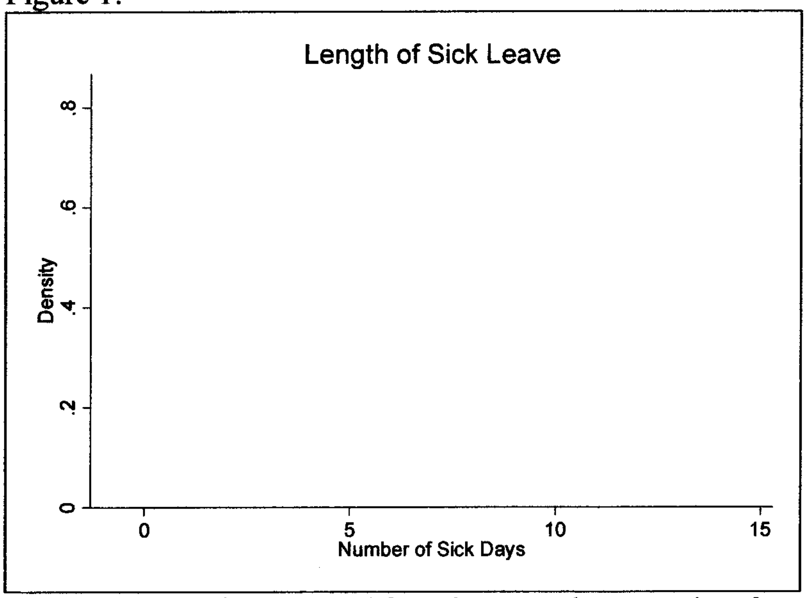
By combining CCHS Cycle 2.1 and Cycle 3.1 this refines the comparable questions I use in my analysis to sixteen. [4] Of these questions, none of the variables I consider are continuous, because nearly all the continuous variables in the public use micro data files have been categorized. In fact, I redefine the endogenous and all exogenous categorical variables as dummy variables, which simplifies the specification and interpretation of the model.[5] By defining dummies for each of the categorical variables the mean of each variable now equals the proportion of individuals who possess a particular characteristic. Also, using dummies makes defining the reference individual much easier because they simply equal the base group (i.e. the category which I drop when creating the dummies from a particular characteristic). I define my hurdle model such that the exogenous variables include dummies for information pertaining to the respondents': age, sex, province of residence, chronic condition, employment status[6], citizenship, cultural origin, highest level of education attained, personal income, proportion of household income, number of adults living in a household, presence of children below the age of 12 and single parents. The endogenous variable corresponds to the number of disability days an

4 While some variables may have been derived slightly differently between the two cycles, the variables chosen for the purposes of the analysis are consistent from one cycle to another.
5 Further details as to the exact construction of all the exogenous and endogenous variables can be found in the Data Appendix.
6 Only individuals who are employed part-time or full-time during the reference period are included in the sample, since unemployed individuals do not require sick leave.

individual takes due to sickness or injury.[7] Therefore, I use this variable to determine the number of sick days an individual takes off work because the other variables of interest were not included in the public use micro data files.

To construct the endogenous variables for my two-part decision model, I start with the categorical variable, *sickdays*, which can take any finite value between 0 and 14, representing the number of sick days in the two week period before the survey (Statistics Canada, 2005c, 2005d) as shown in Figure 1.[8] The figure indicates a very large peak at 0 and two smaller ones at 1 and 14. This is due to the fact that the maximum number of disability days an individual can take off sick is restricted to 14 days.

Figure 1:



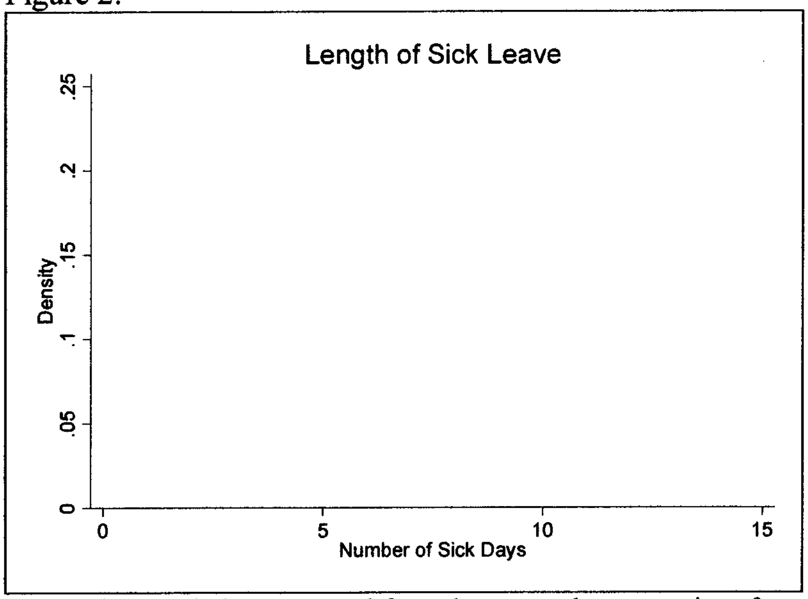
Notes: The sample is constructed from the repeated cross-sections from CCHS and uses working individuals from both sexes between 25 and 64 years of age. N = 99,772.

7 This variable was derived by combining the answers from two other derived variables regarding the number of days an individual spent in bed or cutting down on things they normally do because of the illness or injury during the reference period. This variable does not include time spent in bed or cutting down on tasks because of some emotional/mental health issues or drug or alcohol related illnesses.

8 Technically an individual working a typical work week would not be able to take more than 10 days sick leave over the 14 day period. Regardless, I consider sick days up to 14 days for the purpose of this paper and by doing so I account for those individuals who work 6 to 7 days a week.

I define a new dummy variable to model the binary choice in part one of the hurdle model, whether or not an individual takes a sick day. The dummy dependant variable for the probit decision, *sick*, takes the value 1 if an individual reported at least one sick day and 0 otherwise. I then restrict the categorical dependent variable, *sickdays*, for the second stage of the model to include only those values between 1 and 14. Therefore, the second stage corresponds to a count data model which is represented by a zero-truncated negative binomial model.[9] Figure 2, provides a histogram for the distribution of *sickdays*. Here it is even more obvious that the data is bimodal (i.e. has two peaks).

Figure 2:



Notes: The sample is constructed from the repeated cross-sections from CCHS and uses working individuals from both sexes between 25 and 64 years of age. N = 15,655.

The data set is then restricted to non-missing observations for all the variables which I include in the model. Furthermore, I restrict the sample to individuals 25 to 64 years of age. I drop persons aged 12 to 24 and 65 plus from the sample. While a more standard approach of restricting the sample to workers only consists of including individuals between 18 to 65 years of age and dropping students, here this is not possible. Mostly

9 The theory pertaining to why two dependant variables are necessary is given in the following section.

because the data is grouped in such a way that I am not able to capture individuals aged 65 without including individuals aged 65 to 69. Also, in terms of the compatibility of the two surveys, if I chose to include individuals aged 18 and up I would have to include the group of individuals aged 15 to 19. By doing so, it would be difficult to differentiate students because the data set does not allow me to isolate the month in which the respondents are interviewed. I also drop unemployed individuals from the sample, since it is assumed that these individuals do not take any sick days. Lastly, I drop all values for the proportion of household income the respondent earns which exceed 1, which amounts to 644 observations. Therefore, by restricting the sample accordingly, approximately 37% of all observations remain.[10] This represents a loss of 166,521 observations and another 84,117 observations in step two of the model.

Table 1, provides a list of the descriptive statistics for each of the variables or groups of variables I include in my model, for the whole sample and by gender.[11]

---

10 Initially my subsample consisted of 134,072 observations from Cycle 2.1 and 132,221 observations from Cycle 3.1. for a total of 266,293 observations.

11 Please refer to the Data Appendix for information about how this variable is derived.

Table 1: Descriptive Statistics

Variable	Whole sample		Females		Males	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Dependant variables						
Sick	15.7	36.4	18.7	39.0	12.7	33.3
Sick days	4.7	4.3	4.6	4.2	4.8	4.5
Independent variables						
Female	49.5	50.0				
Age 25 to 29	11.5	31.9	12.4	33.0	10.6	30.8
Age 30 to 34	14.4	35.1	14.8	35.5	14.0	34.7
Age 40 to 44	15.1	35.8	14.3	35.0	15.9	36.5
Age 45 to 49	13.2	33.9	13.6	34.3	12.9	33.5
Age 50 to 54	13.8	34.5	14.3	35.0	13.4	34.0
Age 55 to 59	11.1	31.4	10.9	31.2	11.2	31.6
Age 60 to 64	6.2	24.2	5.4	22.6	7.0	25.6
Part-time worker	12.3	32.8	19.4	39.6	5.3	22.3
Immigrant	12.6	33.2	12.0	32.5	13.1	33.7
Visible minority	9.5	29.4	9.3	29.0	9.8	29.7
Some high school	11.2	31.5	8.9	28.5	13.4	34.1
High school graduate	16.6	37.2	16.7	37.3	16.6	37.2
Some post-secondary	6.6	24.8	6.6	24.9	6.5	24.7
Chronic condition	68.5	46.5	73.6	44.1	63.4	48.2
Children below the age of 12	29.4	45.5	30.1	45.9	28.7	45.2
Single	30.5	46.0	32.6	46.9	28.4	45.1
Single parent	3.8	19.1	6.3	24.2	1.4	11.6
Income between \$0 and \$14,999	11.0	31.3	17.3	37.8	4.9	21.6
Income between \$15,000 and \$29,999	22.4	41.7	29.4	45.6	15.6	36.3
Income between \$50,000 and \$79,999	24.7	43.1	17.3	37.8	31.9	46.6
Income \$80,000 plus	9.8	29.8	4.1	19.9	15.4	36.1
Proportion of household income	72.9	29.2	64.1	31.6	81.5	23.7
Number of observations	99,772		49,343		50,429	

Notes: The descriptive statistics represent percentages, except for the variable sick days. The descriptive statistics for this variable represent real numbers. Furthermore, sick days are only observed for individuals who stay away from the office the sample sizes are as follows: n=15,655 for the whole sample, n=9,236 for females and n=6,419 for males.

After combining the two cycles and defining the variables in the model, 99,772 observations remain in the full model. Of these remaining observations, 49% or 49,373 are females and 51% or 50,429 are male. The working population consists of approximately 12% part-time workers and 88% full-time workers. Also, nearly 30% of individuals have a child below the age of 12 and almost 4% of individuals are single parents. However, given that my analysis focuses on sick leave of the 99,772 individuals in the sample approximately only 16% take any sick leave (also shown in Figure 1). Furthermore, of these individuals who take sick leave, on average individuals take approximately 5 sick days. Therefore, the dummy dependent variable for stage two of the model, *sickdays* includes only those individuals for which *sick* equals 1. This is reflective by the reduced number of observations in the sample, from 99,772 to 15,655. Given the increased restrictions in this portion of the model, of those remaining, 9,236 are female and 6,419 are male.

MODEL

In the Data section I define the categorical variable *sickdays* which I use to construct my dependent variables. In order to model a count data set with a large number of zeros, Cameron and Trivedi (2009, 670) suggest using a two part model. While there are a number of count models which could be used to model this data, because of the underlying distribution of the non-zero absences in Figure 2 (Data), this can lead to overdispersion and biased results. Therefore, the authors suggest a different technique to modelling count data. They suggest partitioning the observations into two data sets, one which models the zero observations and one that models only non-zero counts (Cameron and Trivedi, 680). Modelling a data in this respect, is also ideal when there are two variables of interests and the zero-observations follow a sequential decision process (Jones, 2000), as is the case here. Therefore, I assume that an individual's decision to take a given number of sick days is preceded by their decision to take a sick day and not vice versa.[12]

Consequently, I choose to model my data using a two-part model or hurdle model like Vistnes (1997) does. By doing so, I can use this model to investigate both the likelihood that an individual takes a sick day and the number of sick days they take conditional on them having taken a positive number of sick days. This model therefore, consists of two

12 I ignore the possibility that an individual has been instructed by a medical professional to take a given number of sick days, which then influences their decision to take a sick day.

sub-models, one to model each of the decisions. The first part consists of a binary decision to model inclusion, followed by a second decision which models the number of counts.

Unlike Vistnes (1997), I have chosen to model the decision whether or not an individual takes a sick day using a probit model, instead of a logit model. Therefore, I assume the errors are normally distributed and uncorrelated. This binary choice involves grouping all zero responses and all positive responses as follows:

$$y_1 = \begin{cases} 1 & \text{if } \; y^* > 0 \\ 0 & \text{if } \; y^* \leq 0 \end{cases}, \tag{1}$$

where y_1 represents the variable *sick* and y^* represents the unobserved or latent variable which is a measure of sickness level.

For part two of the hurdle model, like Vistnes (1997) I choose to run a zero-truncated negative binomial for three reasons. First and more importantly, because the alternative would be to use a zero-truncated Poisson model, which requires the assumption of equidispersion (equality of the mean and variance), which here I am able to relax. Subsequently, I do not have to worry as much about the results being overdispersed and or the estimates being inconsistent. Secondly, because it is not necessary to select different variables from one part of the model to the next like in some sample selection models (i.e. Heckman correction). This allows me to compare my results from part one and part two and determine whether or not the factors which contribute to the first decision are the same or different than those in the second decision. Third, it is possible to interpret the results of the zero-truncated negative binomial model as spells. While this is one of Vistnes (1997) principal motivations for choosing this model, it is important to recall that she had data for her dependent variable over a one year period. I, on the other hand, assume only one spell because my reference period is restricted to 14 days.

RESULTS

Table 2 presents the marginal effects from the hurdle model for the entire sample.[13] Recall that the hurdle model I define in the previous section consists of a probit regression in part one and a zero-truncated negative binomial regression in part two. Therefore, the results in part one (column 1, 3 and 5 of Table 2) examine the likelihood that an individual takes a sick day and the second part (column 2, 4 and 6 of Table 2) determines the number of days an individual takes, provided they took a sick day at all. Hence, the interpretation of the marginal effects is slightly different from one part to the other. The marginal effects for the probit model represent the change in the probability that an individual takes a sick day, while the marginal effects for the zero-truncated negative binomial model represent the change in the number of days an individual takes off, given they take a sick day. In addition, for each of the columns in Table 2, I include provincial and yearly fixed effects.

13 I do not include any of the parameter estimates from the hurdle model because they are meaningless when interpreting a two-step model consisting of a probit model and a zero-truncated negative binomial model. What matters is the interpretation of the marginal effects which is similar to that of the coefficients for the linear regression model.

Table 2: Hurdle Model – Combined sample, females and males for 25-64 year old workers (marginal effects reported)

	Whole sample		Females		Males	
	Column 1: Probit model	Column 2: Zero-truncated negative binomial model	Column 3: Probit model	Column 4: Zero-truncated negative binomial model	Column 5: Probit model	Column 6: Zero-truncated negative binomial model
VARIABLES	sick	sickdays	sick	sickdays	sick	sickdays
female	0.0312***	-0.202***				
	[0.00212]	[0.0514]				
age2529	0.00777**	-0.0977	0.0147***	-0.215*	0.00242	0.0711
	[0.00310]	[0.0924]	[0.00512]	[0.118]	[0.00474]	[0.137]
age3034	0.00219	0.00375	0.00152	0.0575	0.00356	-0.0821
	[0.00283]	[0.0887]	[0.00467]	[0.118]	[0.00433]	[0.121]
age4044	-0.00221	0.258***	-0.00149	0.257**	-0.00328	0.260**
	[0.00277]	[0.0930]	[0.00472]	[0.125]	[0.00410]	[0.127]
age4549	-0.0105***	0.441***	-0.00848*	0.329**	-0.0146***	0.550***
	[0.00288]	[0.103]	[0.00492]	[0.133]	[0.00431]	[0.148]
age5054	-0.0154***	0.567***	-0.0144***	0.505***	-0.0203***	0.622***
	[0.00287]	[0.108]	[0.00491]	[0.140]	[0.00432]	[0.154]
age5559	-0.0213***	0.700***	-0.0269***	0.426***	-0.0219***	0.986***
	[0.00299]	[0.118]	[0.00507]	[0.150]	[0.00451]	[0.175]
age6064	-0.0344***	0.679***	-0.0426***	0.394**	-0.0369***	0.932***
	[0.00328]	[0.149]	[0.00571]	[0.192]	[0.00483]	[0.213]
ptjob	0.00544**	-0.0446	0.00622*	0.0153	0.0113*	-0.213*
	[0.00260]	[0.0711]	[0.00362]	[0.0844]	[0.00579]	[0.123]
immigrant	-0.0189***	-0.0089	-0.0237***	-0.0974	-0.0191***	0.0931
	[0.00242]	[0.0842]	[0.00406]	[0.109]	[0.00371]	[0.120]
minority	-0.0117***	0.0265	-0.0156***	0.0672	-0.0118***	-0.028
	[0.00277]	[0.0932]	[0.00459]	[0.126]	[0.00428]	[0.126]
somehs	-0.0203***	0.485***	-0.0246***	0.353***	-0.0212***	0.561***
	[0.00235]	[0.0935]	[0.00423]	[0.130]	[0.00332]	[0.125]
hsgrad	-0.0160***	0.158**	-0.0213***	0.102	-0.0141***	0.256**
	[0.00200]	[0.0717]	[0.00331]	[0.0948]	[0.00308]	[0.103]
someps	0.00942***	-0.0362	0.0132**	-0.0622	0.00814*	-0.0111
	[0.00322]	[0.0895]	[0.00537]	[0.116]	[0.00489]	[0.127]
chronic	0.107***	0.540***	0.137***	0.442***	0.101***	0.546***
	[0.00327]	[0.0764]	[0.00503]	[0.105]	[0.00444]	[0.0982]
child	-0.00391*	-0.118*	-0.00297	-0.145	-0.0043	-0.0539
	[0.00215]	[0.0674]	[0.00371]	[0.0925]	[0.00320]	[0.0905]
single	0.00672***	-0.00958	0.00948**	-0.101	0.00594*	0.108
	[0.00250]	[0.0733]	[0.00452]	[0.105]	[0.00356]	[0.0956]
singleparent	0.00329	-0.0533	0.00236	-0.00788	-0.00281	-0.449*
	[0.00458]	[0.130]	[0.00647]	[0.154]	[0.0105]	[0.263]
inc014	0.0141***	0.360***	0.0167***	0.342***	0.0172***	0.536***
	[0.00331]	[0.0997]	[0.00491]	[0.122]	[0.00645]	[0.179]
inc1529	0.003	0.186***	-0.00177	0.141	0.0124***	0.276**
	[0.00220]	[0.0702]	[0.00333]	[0.0874]	[0.00394]	[0.110]
inc5079	7.10E-05	-0.138**	0.000492	-0.121	0.00175	-0.106
	[0.00212]	[0.0652]	[0.00387]	[0.0932]	[0.00303]	[0.0831]
inc80plus	-0.00703**	-0.233**	-0.00629	-0.600***	-0.00286	-0.00628
	[0.00302]	[0.0957]	[0.00677]	[0.154]	[0.00404]	[0.114]
propinc	0.00676*	0.309**	0.0161**	0.465***	-0.00886	0.126
	[0.00410]	[0.124]	[0.00712]	[0.173]	[0.00644]	[0.172]
Province fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	99772	15655	49343	9236	50429	6419

Notes: The values in parentheses are robust standard errors. The significant marginal effects can be interpreted as follows: *** p<0.01, ** p<0.05, * p<0.1.

To simplify the interpretation of the results in Table 2, I will now define my reference individual. The reference individual for the combined analysis is a Canadian born Caucasian male, between the age of 35 and 39 (inclusive), who resides in Ontario and lives with at least one other adult, has no children, is a post-secondary graduate who works full-time and earns between $30,000 and $49,999 annually (from all sources) which represents almost 62% of household income[14] (from all sources) and is surveyed in 2003. The same applies for the gender specific regressions. Therefore, the predicted probability that the reference individual in the combined sample takes at least one day off is about 8%.[15] The predicted number of sick days that this individual takes, given they take a day off sick is about 2.3 days. It is important when considering the magnitude of the marginal effects in Table 2, that all the results are compared to these predicted values.

From Table 2, given α=5%, the marginal effects in columns 1 and 2 suggest that a number of variables are statistically significant and that the significant variables do not vary much from one part of the model to the other. Column 1 indicates that women are about 3% more likely to take a sick day than men, but that they take about 1/5 of a day less leave than men do when they are sick (column 2). This suggests that women are more likely to take isolated sick days, while men are more likely to take longer bouts of sick leave. Though, the motivation behind this is not clear, my findings prove somewhat contradictory compared to Allebeck and Mastekaasa (2004) findings that women tend to take more sick days than men.

Looking at age, in Table 2, column 1 indicates that for individuals between the ages of 45 to 64 years, age is negatively correlated with taking a sick day. In fact, the magnitude of the probability than an individual takes a sick day becomes increasingly negative with age, from about 1% to 3%. While the literature does indicate that as people age they are

14 This value is calculated by dividing median income level of the respondent by the median household income. I also compute this value for both women and men separately and find that the value is unchanged.
15 For simplicity, I report most of the predicted probabilities to the nearest whole numbers and the predicted number of sick days to the nearest tenth or as fractions for the combined sample and gender specific regressions.

more likely to take sick days (Allebeck and Mastekaasa, 2004; Vistnes, 1997), here the results imply that people who remain in the labour force are healthier. Therefore, those who are in weaker health would select to retire or drop out of the workforce completely. Another possible explanation is that those individuals who are in poorer health may choose to only work part-time which could account for the increase in the probability of those individuals who take a sick day by about 0.5% (in column 1). In column 2, however the signs are reversed, here the magnitudes of the marginal effects increase with age. This suggests that for individuals, this time aged 40 to 64, that the predicted number of sick days an individual takes increases from about ¼ to about 7/10 of a day more sick leave. Furthermore, people who suffer from chronic illnesses are 11% more likely to take sick leave (column 1) and take about ½ day more sick leave (column 2), than healthy individuals. Combining these two findings, the results suggest that as the population ages the probability of developing a chronic illness increases and is consistent with the literature.

Looking at the dummy variables with respect to income in Table 2, for columns 1 and 2 the magnitudes of the marginal effects become increasingly smaller and eventually negative as an individual's income level increases. Although, all four income levels are statistically significant in part two of the model, only the extreme income levels are statistically significant in part one. This suggests that as income increases people are less likely to take a sick day and less likely to a long sick leave if they are ill. This negative correlation between income and sick leave is also consistent with the literature found in Allebeck and Masterkaasa (2004) and Marshall (2006). The sign of marginal effects for the variable *propinc*, however, contradicts the findings above and consequently the literature too, but the variable is only significant at the 5% level in part two of the model. Table 2, also indicates that almost all the marginal effects for the education dummies variables are statistically significant in columns 1 and 2. Comparing the magnitudes and signs of the marginal effects of education in column 1, there appears to be a positive correlation. As the education level of the respondent increases, the probability that they take a sick day increases from -2% for non-high school graduates to 1% for individuals with some post-secondary education. However, the opposite correlation is observed in column 2 and here only two of the three education dummies are significant. The results in column 2 indicate that the more education an individual retains the shorter their time off sick will be. The latter results are consistent with those observed for the income dummies, which emphasizes the positive correlation between education and income.

Coming back to the results for part-time workers, the marginal effects for immigrants and visible minorities are only significant in probit model, except this time the signs of the marginal effects are negative. In other words, individuals who are either a part of a visible minority or who have immigrated to Canada are less likely to take a sick day by about 1% and 2%, respectively. Perhaps this suggests that visible minorities and immigrants are healthier than Caucasians.

I will now focus on the results for children, and single parents in Table 2. For columns 1 and 2, the marginal effects for the presence of children are only statistically significant at the 10% level. In both cases, the signs of the marginal effects are negative which point to a negative correlation between the presence of children and sick leave. Therefore, the results imply that if an individual has a child below the age of 12 they are equally likely to take a day off sick, as someone without any children at all or without children in this age category. Furthermore, parents who have children below the age of 12, tend to take about 1/10 of day less sick leave compared with other individuals. These results are rather ambiguous and given that at the 5% level of significance parents behave in the same manner as other individuals with regards to sick leave, which echoes those of both Vistnes (1997) and Allebeck and Masterkaasa (2004). The marginal effects for single parents are not significant in either column which suggests that single parents and couples behave similarly when confronted with how much sick leave to take. Lastly, the results in Table 2 indicate that single individuals are about 1% more likely to take a sick day (column 1), but are equally as likely to take the same number of sick days as couples (column 2). This suggests they may be more willing to take sick leave when they need it because they only have to earn enough incomes for themselves.

The remaining four columns in Table 2 represent the marginal effects for females (columns 3 and 4) and males (column 5 and 6). Similarly, columns 3 and 5 represents the marginal effects from the probit model and columns 4 and 6 represent the results of the zero-truncated negative binomial model. However, before I proceed to compare these results with those in columns 1 and 2, it is important to recall that other than gender, the reference individual unchanged. Therefore, the predicted probability that the female reference individual takes a sick day is approximately 11%, in column 3. The predicted number of sick days she takes, given she takes at least one day off is about 2.4 sick days, for column 4. Compare this to the male reference individual, whose predicted probability of taking a sick day is approximately, 9% for column 5 and the predicted number of sick days he takes given he takes a sick day is about 2.1 days. Again, it is important when considering the magnitude of the marginal effects in columns 3 to 6, that all the marginal effects are interpreted with respect to these predicted values.

Comparing columns 3 to 6, to columns 1 and 2 of Table 2, the results are fairly consistent with the ones described above. In addition, the sign of the marginal effects from one column to the next appear unchanged. Regardless, there are a few differences that are noted when comparing the results from the columns 3 to 6. Most notable are the differences in the significant variables for the age dummies, *child*, *singleparent* and *single*.

Comparing age dummies across columns 3 to 6, the results for women and men are fairly consistent. At the 5% level of significance, the age results for the zero-truncated negative binomial regressions (columns 2, 4 and 6) are identical in terms of sign and significance. However, the magnitudes of the marginal effects are larger for men than the combined sample and smaller for women than in the combined sample. Comparing the results, for the probit models the findings are less consistent. While the results for men are identical to the combined sample, at the 5% level of significance between the ages of 45 to 64, the dummy variable for individuals aged 25 to 29 is not significant for males. Conversely, for the female probit regression the dummy variable for women aged 25 to 29 is statistically significant at the 5% level and reflects a 1% increase in the probability that a woman takes a day off sick. This finding is consistent with women's child bearing years as per literature review done by Allebeck and Mastekaasa (2004). The only difference between the significant ages of women compared to the combined sample is that women between the age of 45 and 49 years of age insignificant at the 5% level, meaning there is no change in the probability that they take a sick day.

Now considering the results for children in columns 3 to 6, the significant results differ considerably. For both men and women, children have no statistical significance in either stage of the regression, not even at the 10% level like for the combine sample. Regardless, the signs of the marginal effects in columns 3 to 6 are consistent with those observed in columns 1 and 2, although the magnitudes more closely reflect the results of the female regressions. However, because the presence of children below the age of 12 is not significant in columns 3 to 6, this suggests that once again the sick leave of parents with children in this age category is similar to that of other individuals.

Lastly, I will focus on the significance of single parents and single individuals across genders. As is readily seen from columns 3 and 4 of Table 2, the marginal effect for single mothers is not statistically significant in the model meaning the likelihood that they take sick leave and the number of days they take is the same as mothers from two parent households. However, at the 10% level of significance single fathers take about 2/5 of a day less sick leave (column 6). Conversely, the marginal effect for single males is only statistically significant at the 10% level in part one of the model (column 5). This suggests that single males are about 0.5% more likely to take a sick day compared to married or committed men.[16] Similarly, the marginal effect for single women is only significant in the first part of the model, but this time the results are significant at the 5% level of significance. Therefore, single women are about 1% more likely to take a sick day, but there is no difference between the number of days single women take off compared with married or committed women.

16 There are 682 single fathers included in part one of the hurdle model which consists of 50,429 males and 83 single fathers in part two of the hurdle model which consists of 6,419 males.

DISCUSSION

One of the major challenges that presents itself when using a public use micro data file is that the data may be re-coded such that none of the responses are continuous. In fact, all the variables I chose for my model were defined as either dummy or categorical variables. As mentioned earlier, some difficulty arose when I was trying to isolate for the working population because I was unable to include individuals aged 65 in the sample. Also, using income categories to derive the variable, *propinc*, (i.e. the proportion of household income the respondent earns) proved to be a bit challenging and resulted in some observations having to be dropped because the proportions exceeded 1.[17]

Another potential problem that arises has to do with endogeneity bias. In performing my variable selection I chose to include a health related indicator like Vistnes (1997) advocated. The only problem is that by doing so this may have caused some of the variables to become correlated with each other (i.e. multicollinearity) or worse to become correlated with the error term, resulting in an endogeneity problem. However, to reduce the amount of self selection with respect to this variable I chose to include a dummy variable for chronic illness instead of a health indicator for self-reported health status. Furthermore, before including the variable for chronic illness I chose to generate summary statistics for the entire sample by age. My findings are as the theory predicts, as people age their health deteriorates. Furthermore, I was unable to control for the type of job or industry in which an individual works because the data was not provided in the data set. Therefore, I was unable to compare my findings with those of van Poppel et al. (2002) regarding ergonomic load, as I had hoped. Likewise, I was unable to control for health insurance, job tenure or unionized workers, like Vistnes (1997) and Marshall

17 An explanation of how this variable is constructed is provided in the Data Appendix.

(2006) do. Thus, by restricting the model accordingly, it may now suffer from omitted variable bias, which is another form of endogeneity bias.

An additional problem which occurs has to do with the survey design, since no survey is completely random, although it is often assumed to be. In particular, the CCHS is constructed from a stratified sample, to include over 120 health regions from each of the provinces and territories (as mentioned in the Data section). The problem that arises from sampling schemes is that the error terms are no longer identically distributed. Furthermore, should any correlations exist across individuals living in the same geographical area; the errors will no longer satisfy the assumption of independence, which can cause clustering issues.

The most significant limitation here has to do with self-selection or self-reported bias. This problem results in the estimators being either upwards or downwards biased and generally results from the respondent desire not to answer truthfully or they simply forget the appropriate response.[18] Both of these issues are likely to cause some measurement error (not in the classical sense) in the number of sick days an individual reports, though they could affect other variables too. Similarly, this type of problem can also cause an individual to answer "not stated" or "not applicable" to a survey question. Alternatively, a respondent may answer a question using one of these two responses if they refuse to answer the question or simply do not know the answer and "refusal" or "don't know" are not specified. Either way, each time a respondent selects one of these four responses to a survey question another missing observation is generated. This combined with my restrictions results in considerable attrition in the sample. In fact, the hurdle model for

18 Although, based on the article by van Poppel et al. (2002), it is likely that the variable sick leave did suffer from recall bias as much as other variables because the reference period was restricted to only two weeks and not a year.

the whole sample results in 166,521 observations being dropped in probit model and another 84,117 observations being dropped in the zero-truncated negative binomial model.

Finally, in an attempt to determine whether or not my results were robust, I ran a number of sensitivity analyses on my results. In fact, I ran four different hurdle models using various variable specifications and three time sensitivity regressions, including all of the variables in my model. I not only ran the sensitivity analyses for the combined sample but by gender too.[19]

The four different variable specifications I tried exclude: chronic illness; children below the age of 12 and single parents, simultaneously; education dummies and income dummies. The results were fairly similar to those observed in Table 2, in terms of sign and variable trends, though some notable differences are observed. Most notable are the differences in the number of significant variables from one specification to another and the magnitudes of the marginal effects.[20] In fact, the results for the zero-truncated negative binomial and for all four specifications seem to vary only in terms of the magnitudes of the marginal effects (Tables 4, 5, 6). However, the results for the probit regression model were not as consistent. First, for the male sample the sign of the variable *propinc*, changes from positive to negative, although the variable remains insignificant (Table 6). Secondly, the variable *child* becomes statistically significant at the 5% level in all three samples when the variable for chronic illness is excluded from the model (it is only significant at the 10% level for women). Thirdly, when I exclude the education dummies and the variables child and singleparent, simultaneously from the

---

19 The results from each of the sensitivity analyses are provided in Tables 4 to 9 in the Appendix.

20 It is important to note that the numbers of observations, in each of the four regressions, are not all equal. The same thing goes for the time sensitivity results.

model, the variable *single* becomes statistically significant at the 5% level for men, like in the other two regressions. Lastly, the specification which differed the most overall was the one for which I excluded the variable *chronic*. This suggests that this variable may be correlated with either some of the other independent variables or the error term, which may result in endogeneity bias.

Moving on to the results from the time sensitivity analyses, here I restricted the maximum number of sick days in the reference period to 7 days, 10 days and 13 days. I chose to run three such analyses to see if the results varied much by excluding the spike in observations at 14 sickdays (Figure 1 and 2 in the Data section). I chose to restrict the variable *sickdays* in the first specification to one week, similar to short leave in van Poppel et al. (2002). I then chose to restrict the second specification for *sickdays* to 10 days to account for the maximum number of sick days an individual working a 5 day work week could take off. Finally, for the third specification I chose to restrict the variable *sickdays* to 13 days, thereby excluding the spike at 14 days.

Unlike the results for the variable specifications, here the results were less consistent with those in Table 2. Although, this time the results from the probit regressions were more consistent for all three samples in terms of signs and variable trends (Tables 7, 8, 9). However, the income variables become less significant for all three specifications. Like before, the magnitude of the marginal effects are different from those observed in Table 2 for both the whole sample regressions and gender specific ones. However, these differences are minor compared to those observed for the zero-truncated negative binomial model for each of the three specifications and for all three samples. Most notable is that in all three of the models the age dummies are all no longer significant.

For all three specifications for females, nearly all the marginal effects for age are now negative (Table 8). Similarly, the marginal effects are also mostly negative in two out of the three specifications for the whole sample (Table 7), which suggests that the results are highly sensitive to the number of sick days. While the income and education trends are maintained in each set of sample results, the variables vary in terms of significance and a number of the variables are no longer significant. Furthermore, the specification that varies the most for both parts of the hurdle model is when *sickdays* is defined as being between 1 and 7 days, inclusive. Therefore, this suggests the results are biased.

In fact the data itself is biased by way of construction. First of all, the variable *sickdays* is restricted to a two-week period only, which causes the data to be censored above. Due to this limited reference period, the number of sick days I am observing more than likely contributes only to one spell of sick leave and not the average length of a particular spell over a longer period. This is perhaps why Vistnes (1997) chose to perform her analysis using data for sick leave over a one year period. In addition, because I cannot control for the month or season in which the respondents were interviewed it is possible that those interviewed during the winter months were more likely to be sick than those during the summer months. As a result, both of these factors may contribute to the time sensitivity of the results.

CONCLUSION

While constructing a two-step hurdle model does have its advantages, in this case it allows me to isolate the probability that an individual takes a sick day, from the predicted number of sick days they take, conditional on them taking a sick day. It also proves to be quite advantageous when dealing with the large number of zero-observations in the sample when going from one step to the other. As a result, I was able to examine the determinants of sick leave and length of leave simultaneous, and compare them accordingly. It turns out that age, income level, education level and chronic illness were all key indicators in predicting both the probability of taking a sick day and determining the number of sick days an individual takes in each of the samples. In addition, gender also proves to be significant in the combined model as women are about 3% more likely to take a sick day than men and about 1/5 of a day less leave than men do when they are sick. Furthermore, some notable differences are observed in the gender specific regressions.

My findings regarding age suggest that regardless of gender, as individuals' age they are less likely to take a sick day, but more likely to take a longer leave. In fact, individuals between the ages of 45 and 64 are more likely to take longer leave if they do take a sick day.

Furthermore, my findings for income and education suggest that the higher the income level or education level attainted by the respondent the less likely they are to take a sick day and they take short bouts of sick leave when they are sick.

Conversely, my results regarding the presence of children suggest that overall parents take the same amount of sick leave as other individuals for the combined sample and by gender, though their impact was not completely insignificant. For the combined sample, at the 10% level parents had the same probability of taking a sick day, but parents were more likely to take about 1/10 of a day less sick leave. In addition, for single fathers at the 10% level, they were more likely to take about half a day less sick leave. Though, overall couples and single parents are equally likely to take a sick day and tend to take the same amount of sick leave.

Finally, the results of the sensitivity analyses suggest that the model is both sensitive to variable specification and number of sick days. Specifically, the exclusion of the variable *chronic* for the whole sample suggests that individuals with children behave in the same manner as other individuals regarding the number of sick days they take, but are less likely to take a sick day. Likewise, limiting the number of sick days to 7, the results in part two for all three samples are otherwise insignificant for all variables of interest.

While there is already great deal of literature already available regarding sick leave perhaps more needs to be done to specifically reflect the role of children on single parents or on adults with ailing parents. Moreover, it would be interesting to see if there are any measurable differences in the length of sick leave due to ailing parents versus children, or ailing parents versus children with severe disabilities or diseases.

DATA APPENDIX

The data sources for this particular analysis are from the Canadian Community Health Survey Public Use Micro-Data Files Cycle 2.1 and Cycle 3.1. Therefore, when combining any cross-sections together it is necessary to first compare the questionnaires, codebooks and derived variables before constructing any variables or models (Thomas and Wannell, 2009). Once, the responses are compared and the variables selected the data construction can begin. Subsequently, all non-responses are first dropped from the sample before any of the variables in the model are defined. Therefore, it is necessary to drop all responses for “not stated”, “refusal”, “not applicable”, “don’t know” or “other”.

In fact, these options exist for most survey questions, except for age, sex, province of residence, children below the age of 5 and children between the ages of 6 and 11. Consequently, the repeated cross-section includes only those individuals who are working members of the labour force between 25 and 64 years of age. However, these questions were originally intended for all respondents between 15 and 74 years of age, inclusively. By restricting the age category, I assume that only permanent full-time (and part-time) workers remain; thereby eliminating students (who work only full-time during the summer, a co-operative work term or internship) or seasonal workers from the sample.

One limitation that often arises when using a PUMF is that the data is frequently derived from the master file, as is the case for the age variable. The respondent's age is no longer given, only the category in which their age falls. Most of the categories included individuals within a five-year age group, which I redefine as dummy variables for all but one age category, age 35 to 39. I use a similar strategy when restructuring the dummy variables for province of residence, highest level of education attained, personal income and household income. For each of these variables the reference individual is from Ontario, has a post-secondary degree, earns between $30,000 and $49,999 and has a household income between $50,000 and $79,999.

The only one new categorical variable that I derive for the purpose of this model is *propinc*. This variable represents the proportion of total household income earned by the respondent and is derived from both the variables for personal and household income. The first step involves generating a new variable which assigns a median value to each of the income ranges for both personal and household income.[21] The variables are given the names *income* and *hincome*, respectively. Lastly, to generate the variable *propinc*, I simply divide the value for *income* by *hincome*, which gives me an estimate for the proportion of household income. Hence, the reference individual earns roughly 62% of the household income (i.e. $40,000/$65,000 * 100).

Similarly, I also derive three new dummy variables for the model namely, *child single* and *singleparent*. The first variable *child* indicates whether or not the respondent has any children below the age of 12. I construct it by assigning a value of 1 if the individual has a child below the age of 5 (*child05*) or between the ages of 6 and 11 (*child611*) (or both) and 0 otherwise. The second variable *single* corresponds to the number of adults living in a given household. The value 1 represents a household with only 1 adult, be it a single parent or an individual living in solitude and 0 represents a household with more than one adult. Since I derive this variable from the survey variable regarding living arrangement, more than one adult can mean an individual living with their partner/spouse, roommates or partner/spouse and children. Finally, the last variable *parent* is an interaction variable which I derive by multiplying the two previous variables together (i.e. *singleparent* = *single*child*). Therefore, if *singleparent* = 1 the individual is a single parent and 0 otherwise. Hence, the reference individual has no children and lives with at least one other adult.

There are also a few variables that are already defined as dummy variables (or were reduced to dummy variables once the non-response answers are dropped) these were

21 The median income values were all rounded to whole numbers (in dollars). In addition, each of the income ranges, including the reference individual's income level was also set to the median value. For instance, the reference individual earns between $30,000 and $49,999, therefore, their median personal income value is $40,000.

simply renamed for ease of interpretation. Such variables include: *female*, *immigrant*, *minority*, *chronic*, *child05* and *child611*.

Lastly, the two dependent variables are both derived from the same variable which represents the number of disability days an individual takes off, as described earlier. Table 3, provides a list of all the variables in the model along with their survey variable names.

Table 3: Variable Specifications

Group Description	Further Description	Variable Name	Survey Variable Name	
			Cycle 2.1	Cycle 3.1
Two-week Disability	Number of sick days an individual takes between 0 and 14 (inclusive)	sickdays	TWDCDDDP	TWDEDDDP
	Individual takes at least one day of sick leave Individual takes no sick leave	sick *	TWDCDDDP	TWDEDDDP
Province of residence	Newfoundland and Labrador Prince-Edward Island Nova Scotia New Brunswick Quebec Ontario Manitoba Alberta Saskatchewan British Colombia North West Territories, Yukon and Nunavut	nfl pei ns nb qc * mb ab sk bc terr	GEOCGPRV	GEOEGPRV
Age categories	Age 25 to 29 Age 30 to 34 Age 35 to 39 Age 40 to 44 Age 45 to 49 Age 50 to 54 Age 55 to 59 Age 60 to 64	age2529 age3034 * age4044 age4549 age5054 age5559 age6064	DHHCGAGE	DHHEGAGE
Sex	Female Male	female *	DHHC_SEX	DHHE_SEX
Employment status	Part-time job Full-time job	ptjob *	LBFCDPFT	LBSEDPFT
Canadian citizenship	Immigrant Canadian citizen	immigrant *	SDCCFIMM	SDCEFIMM
Cultural origin	Visible minority White	minority *	SDCCGRAC	SDCEGCGT
Chronic Condition	Individual has at least one chronic condition Individual has no chronic conditions	chronic *	CCCCF1	CCCEF1
Highest level of education attainted	Some high school High school graduate Some post-secondary education University graduate	somehs hsgrad someps *	EDUCDR04	EDUEDR04
Children	Household with a least one child under 5 years of age No children in this age category	child05 *	DHHCGLE5	DHHEGLE5
	Household with at least one child between 6 and 11 years of age No children in this age category	child611 *	DHHCG611	DHHEG611
	Household with a least one child below the age of 12 No children in this age category	child *	DHHCGLE5 DHHCG611	DHHEGLE5 DHHEG611
Living arrangement	Household with only one adult Household with more than one adult	single *	DHHCGLVG	DHHEGLVG
	single*child	singleparent	DHHCGLVG DHHCGLE5 DHHCG611	DHHEGLVG DHHEGLE5 DHHEG611
Personal Income (from all sources)	$0 - $14,999 $15,000 - $29,999 $30,000 - $49,999 $50, 000 - $79,999 $80,000 +	inc014 inc1529 * inc5079 inc80plus	INCCGPER	INCEGPER
Household Income (from all sources)	$0 - $14,999 $15,000 - $29,999 $30,000 - $49,999 $50, 000 - $79,999 $80,000 +	hinc014 hinc1529 hinc3049 * hinc80plus	INCCGHH	INCEGHH
Proportion of household income	personal income/household income	propinc	INCCGPER INCCGHH	INCEGPER INCEGHH

Note: Reference category is represented by the asterisk (*).

APPENDIX

Table 4: Variable Senitvity Analysis - Whole Sample

	Probit marginal effects				Zero-truncated negative binomial marginal effects			
	No *chronic*	No *child* or *singleparent*	No education dummies	No income dummies	No *chronic*	No *child* or *singleparent*	No education dummies	No income dummies
VARIABLES	sick	sick	sick	sick	sickdays	sickdays	sickdays	sickdays
female	0.0572***	0.0306***	0.0319***	0.0324***	-0.209***	-0.203***	-0.242***	-0.162***
	[0.00302]	[0.00205]	[0.00207]	[0.00200]	[0.0607]	[0.0494]	[0.0524]	[0.0466]
age2529	0.00413	0.00839***	0.00847***	0.00882***	-0.152	-0.0785	-0.114	-0.0484
	[0.00455]	[0.00302]	[0.00295]	[0.00308]	[0.108]	[0.0894]	[0.0950]	[0.0899]
age3034	-0.000313	0.00223	0.00273	0.00248	-0.0155	0.0028	-0.00472	0.0168
	[0.00422]	[0.00277]	[0.00269]	[0.00281]	[0.104]	[0.0860]	[0.0913]	[0.0861]
age4044	-0.000384	-0.00159	-0.00305	-0.00233	0.331***	0.273***	0.284***	0.248***
	[0.00422]	[0.00268]	[0.00261]	[0.00274]	[0.110]	[0.0905]	[0.0961]	[0.0898]
age4549	-0.0106**	-0.00894***	-0.0115***	-0.0106***	0.549***	0.486***	0.480***	0.421***
	[0.00445]	[0.00269]	[0.00271]	[0.00286]	[0.121]	[0.0991]	[0.106]	[0.0988]
age5054	-0.0139***	-0.0135***	-0.0160***	-0.0155***	0.724***	0.629***	0.616***	0.539***
	[0.00450]	[0.00261]	[0.00271]	[0.00285]	[0.126]	[0.103]	[0.111]	[0.103]
age5559	-0.0204***	-0.0193***	-0.0217***	-0.0213***	0.902***	0.764***	0.763***	0.674***
	[0.00475]	[0.00270]	[0.00281]	[0.00296]	[0.139]	[0.114]	[0.123]	[0.114]
age6064	-0.0413***	-0.0323***	-0.0344***	-0.0340***	0.862***	0.744***	0.775***	0.683***
	[0.00539]	[0.00300]	[0.00306]	[0.00323]	[0.176]	[0.145]	[0.156]	[0.145]
ptjob	0.00923**	0.00507**	0.00616**	0.00928***	-0.0399	-0.0535	-0.0567	0.0604
	[0.00392]	[0.00253]	[0.00249]	[0.00250]	[0.0845]	[0.0685]	[0.0729]	[0.0669]
immigrant	-0.0371***	-0.0185***	-0.0167***	-0.0185***	-0.0525	-0.0083	-0.0311	0.00706
	[0.00366]	[0.00236]	[0.00229]	[0.00239]	[0.0978]	[0.0816]	[0.0858]	[0.0821]
minority	-0.0199***	-0.0115***	-0.0112***	-0.0107***	0.0268	0.0216	0.0341	0.0598
	[0.00427]	[0.00271]	[0.00262]	[0.00275]	[0.110]	[0.0903]	[0.0961]	[0.0912]
somehs	-0.0312***	-0.0198***		-0.0180***	0.589***	0.472***		0.594***
	[0.00365]	[0.00228]		[0.00228]	[0.110]	[0.0906]		[0.0926]
hsgrad	-0.0274***	-0.0156***		-0.0149***	0.183**	0.150**		0.206***
	[0.00304]	[0.00194]		[0.00195]	[0.0845]	[0.0693]		[0.0701]
someps	0.0151***	0.00924***		0.0107***	-0.046	-0.0384		0.027
	[0.00479]	[0.00316]		[0.00323]	[0.106]	[0.0867]		[0.0882]
chronic		0.105***	0.103***	0.107***		0.529***	0.565***	0.544***
		[0.00309]	[0.00316]	[0.00313]		[0.0740]	[0.0787]	[0.0741]
child	-0.00940***		-0.0033	-0.00384*	-0.165**		-0.132*	-0.123*
	[0.00324]	0.00797***	[0.00204]	[0.00213]	[0.0791]		[0.0693]	[0.0651]
single	0.00850**	[0.00233]	0.00813***	0.00917***	-0.0189	0.0115	-0.0429	0.102*
	[0.00371]		[0.00240]	[0.00195]	[0.0866]	[0.0675]	[0.0746]	[0.0556]
singleparent	0.00976		0.0021	0.00383	-0.036		-0.0266	-0.0128
	[0.00702]		[0.00430]	[0.00456]	[0.155]		[0.136]	[0.128]
inc014	0.0274***	0.0136***	0.00884***		0.457***	0.341***	0.454***	
	[0.00497]	[0.00323]	[0.00299]		[0.118]	[0.0961]	[0.103]	
inc1529	0.00674**	0.00293	0.00073		0.229***	0.178***	0.237***	
	[0.00334]	[0.00215]	[0.00204]		[0.0828]	[0.0679]	[0.0726]	
inc5079	-0.00184	-8.95E-06	0.00224		-0.175**	-0.137**	-0.173***	
	[0.00320]	[0.00208]	[0.00202]		[0.0767]	[0.0632]	[0.0664]	
inc80plus	-0.0151***	-0.00701**	-0.0033		-0.292***	-0.229**	-0.290***	
	[0.00455]	[0.00295]	[0.00290]		[0.112]	[0.0928]	[0.0965]	
propinc	0.0137**	0.00648	0.00377		0.389***	0.287**	0.367***	
	[0.00622]	[0.00400]	[0.00388]		[0.146]	[0.120]	[0.129]	
Province fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	99842	99772	99772	99772	15665	15655	15655	15655

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Variable Senitvity Analysis - Females

	Probit marginal effects				Zero-truncated negative binomial marginal effects			
	No *chronic*	No *child* or *singleparent*	No education dummies	No income dummies	No *chronic*	No *child* or *singleparent*	No education dummies	No income dummies
VARIABLES	sick	sick	sick	sick	sickdays	sickdays	sickdays	sickdays
age2529	0.0133*	0.0151***	0.0158***	0.0154***	-0.284**	-0.192*	-0.220*	-0.154
	[0.00750]	[0.00505]	[0.00499]	[0.00506]	[0.135]	[0.113]	[0.119]	[0.114]
age3034	-0.00143	0.00151	0.00224	0.00168	0.0502	0.0511	0.0551	0.0724
	[0.00700]	[0.00462]	[0.00454]	[0.00463]	[0.137]	[0.114]	[0.119]	[0.114]
age4044	0.00102	-0.000925	-0.00314	-0.00164	0.314**	0.279**	0.281**	0.226*
	[0.00718]	[0.00461]	[0.00455]	[0.00468]	[0.145]	[0.121]	[0.126]	[0.119]
age4549	-0.0067	-0.00725	-0.0107**	-0.00857*	0.391**	0.391***	0.354***	0.300**
	[0.00756]	[0.00459]	[0.00472]	[0.00488]	[0.153]	[0.126]	[0.134]	[0.127]
age5054	-0.0108	-0.0130***	-0.0162***	-0.0143***	0.615***	0.583***	0.541***	0.476***
	[0.00762]	[0.00446]	[0.00471]	[0.00486]	[0.161]	[0.133]	[0.142]	[0.134]
age5559	-0.0269***	-0.0254***	-0.0285***	-0.0265***	0.533***	0.505***	0.468***	0.412***
	[0.00804]	[0.00460]	[0.00485]	[0.00501]	[0.173]	[0.143]	[0.152]	[0.145]
age6064	-0.0529***	-0.0411***	-0.0439***	-0.0417***	0.485**	0.473**	0.450**	0.416**
	[0.00946]	[0.00529]	[0.00543]	[0.00563]	[0.223]	[0.186]	[0.196]	[0.188]
ptjob	0.0104*	0.00591*	0.00719**	0.00922***	0.0257	0.000601	0.00873	0.0972
	[0.00543]	[0.00355]	[0.00354]	[0.00338]	[0.0982]	[0.0804]	[0.0848]	[0.0776]
immigrant	-0.0467***	-0.0234***	-0.0218***	-0.0232***	-0.136	-0.0925	-0.109	-0.0775
	[0.00618]	[0.00400]	[0.00393]	[0.00401]	[0.125]	[0.105]	[0.109]	[0.106]
minority	-0.0265***	-0.0155***	-0.0154***	-0.0147***	0.0795	0.056	0.0761	0.0912
	[0.00713]	[0.00453]	[0.00442]	[0.00455]	[0.147]	[0.121]	[0.128]	[0.123]
somehs	-0.0373***	-0.0243***		-0.0220***	0.406***	0.345***		0.473***
	[0.00670]	[0.00416]		[0.00410]	[0.149]	[0.126]		[0.127]
hsgrad	-0.0366***	-0.0210***		-0.0204***	0.114	0.0941		0.159*
	[0.00507]	[0.00325]		[0.00321]	[0.110]	[0.0911]		[0.0916]
someps	0.0213***	0.0131**		0.0140***	-0.0847	-0.0654		0.00881
	[0.00789]	[0.00531]		[0.00533]	[0.134]	[0.111]		[0.114]
chronic		0.136***	0.134***	0.136***		0.433***	0.445***	0.452***
		[0.00481]	[0.00494]	[0.00486]		[0.101]	[0.106]	[0.101]
child	-0.00945*		-0.00213	-0.00299	-0.187*		-0.158*	-0.148*
	[0.00558]		[0.00360]	[0.00367]	[0.107]		[0.0931]	[0.0886]
single	0.0149**	0.0101**	0.0116***	0.0165***	-0.119	-0.0908	-0.125	0.0951
	[0.00673]	[0.00428]	[0.00442]	[0.00331]	[0.122]	[0.0968]	[0.106]	[0.0740]
singleparent	0.00883		0.00049	0.00293	0.0126		0.0182	0.0266
	[0.00988]		[0.00618]	[0.00642]	[0.180]		[0.157]	[0.150]
inc014	0.0333***	0.0164***	0.00968**		0.422***	0.322***	0.411***	
	[0.00729]	[0.00484]	[0.00447]		[0.141]	[0.117]	[0.123]	
inc1529	0.000417	-0.00177	-0.00463		0.175*	0.132	0.174**	
	[0.00508]	[0.00329]	[0.00316]		[0.101]	[0.0841]	[0.0881]	
inc5079	-0.00167	0.00044	0.00337		-0.149	-0.119	-0.141	
	[0.00582]	[0.00383]	[0.00377]		[0.108]	[0.0898]	[0.0931]	
inc80plus	-0.016	-0.00636	-0.00235		-0.712***	-0.584***	-0.624***	
	[0.0102]	[0.00669]	[0.00668]		[0.175]	[0.148]	[0.154]	
propinc	0.0292***	0.0162**	0.0125*		0.548***	0.461***	0.497***	
	[0.0106]	[0.00703]	[0.00678]		[0.207]	[0.167]	[0.174]	
Province fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49369	49343	49343	49343	9240	9236	9236	9236

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Variable Senitvity Analysis - Males

	Probit marginal effects				Zero-truncated negative binomial marginal effects			
	No *chronic*	No *child* or *singleparent*	No education dummies	No income dummies	No *chronic*	No *child* or *singleparent*	No education dummies	No income dummies
VARIABLES	sick	sick	sick	sick	sickdays	sickdays	sickdays	sickdays
age2529	-0.00487	0.00337	0.00328	0.00415	0.059	0.0885	0.0451	0.0996
	[0.00680]	[0.00460]	[0.00453]	[0.00475]	[0.164]	[0.134]	[0.143]	[0.138]
age3034	0.000815	0.0037	0.00407	0.00413	-0.12	-0.076	-0.0979	-0.0725
	[0.00626]	[0.00423]	[0.00414]	[0.00434]	[0.145]	[0.119]	[0.127]	[0.121]
age4044	-0.00168	-0.00275	-0.00373	-0.00337	0.350**	0.263**	0.272**	0.268**
	[0.00608]	[0.00399]	[0.00389]	[0.00410]	[0.154]	[0.125]	[0.133]	[0.127]
age4549	-0.0163**	-0.0130***	-0.0151***	-0.0152***	0.714***	0.564***	0.598***	0.550***
	[0.00649]	[0.00406]	[0.00408]	[0.00430]	[0.176]	[0.145]	[0.155]	[0.146]
age5054	-0.0205***	-0.0181***	-0.0203***	-0.0209***	0.829***	0.648***	0.664***	0.600***
	[0.00658]	[0.00394]	[0.00409]	[0.00431]	[0.184]	[0.149]	[0.161]	[0.151]
age5559	-0.0199***	-0.0195***	-0.0221***	-0.0226***	1.302***	1.015***	1.062***	0.967***
	[0.00697]	[0.00409]	[0.00426]	[0.00449]	[0.207]	[0.171]	[0.184]	[0.171]
age6064	-0.0433***	-0.0344***	-0.0369***	-0.0375***	1.219***	0.963***	1.070***	0.942***
	[0.00769]	[0.00440]	[0.00453]	[0.00479]	[0.257]	[0.209]	[0.227]	[0.211]
ptjob	0.0197**	0.0111*	0.0125**	0.0175***	-0.237	-0.212*	-0.249*	-0.0576
	[0.00843]	[0.00566]	[0.00560]	[0.00588]	[0.150]	[0.120]	[0.128]	[0.125]
immigrant	-0.0364***	-0.0187***	-0.0166***	-0.0187***	0.0487	0.0908	0.0579	0.11
	[0.00544]	[0.00362]	[0.00353]	[0.00370]	[0.140]	[0.118]	[0.124]	[0.121]
minority	-0.0198***	-0.0117***	-0.0114***	-0.0103**	-0.0488	-0.031	-0.0246	0.0178
	[0.00639]	[0.00417]	[0.00406]	[0.00431]	[0.151]	[0.124]	[0.132]	[0.127]
somehs	-0.0319***	-0.0206***		-0.0188***	0.713***	0.551***		0.661***
	[0.00499]	[0.00322]		[0.00325]	[0.150]	[0.122]		[0.128]
hsgrad	-0.0234***	-0.0138***		-0.0131***	0.305**	0.240**		0.282***
	[0.00455]	[0.00300]		[0.00304]	[0.123]	[0.100]		[0.103]
someps	0.0124*	0.00794*		0.00996**	0.00259	-0.0149		0.0468
	[0.00709]	[0.00478]		[0.00495]	[0.154]	[0.124]		[0.129]
chronic		0.0989***	0.0970***	0.101***		0.541***	0.594***	0.559***
		[0.00420]	[0.00429]	[0.00429]		[0.0960]	[0.103]	[0.0981]
child	-0.00943**		-0.00379	-0.00540*	-0.0965		-0.0648	-0.0858
	[0.00470]		[0.00305]	[0.00319]	[0.108]		[0.0950]	[0.0896]
single	0.00603	0.00737**	0.00727**	0.00403	0.117	0.107	0.075	0.143*
	[0.00515]	[0.00322]	[0.00342]	[0.00295]	[0.115]	[0.0873]	[0.0987]	[0.0805]
singleparent	-0.00635		-0.00313	-0.00232	-0.542*		-0.438	-0.455*
	[0.0154]		[0.00997]	[0.0106]	[0.317]		[0.285]	[0.259]
inc014	0.0311***	0.0169***	0.0139**		0.688***	0.538***	0.613***	
	[0.00935]	[0.00630]	[0.00602]		[0.214]	[0.176]	[0.189]	
inc1529	0.0197***	0.0122***	0.00978***		0.327**	0.276**	0.345***	
	[0.00569]	[0.00385]	[0.00368]		[0.131]	[0.108]	[0.117]	
inc5079	0.00075	0.00163	0.00395		-0.141	-0.106	-0.157*	
	[0.00443]	[0.00296]	[0.00290]		[0.0998]	[0.0814]	[0.0861]	
inc80plus	-0.00792	-0.00284	0.00118		-0.0299	-0.00373	-0.0891	
	[0.00590]	[0.00394]	[0.00389]		[0.137]	[0.112]	[0.116]	
propinc	-0.0132	-0.00932	-0.0117*		0.202	0.118	0.203	
	[0.00949]	[0.00627]	[0.00611]		[0.206]	[0.168]	[0.179]	
Province fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	50473	50429	50429	50429	6425	6419	6419	6419

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Time Senitvity Analysis - Whole Sample

	Probit marginal effects			Zero-truncated negative binomial marginal effects		
	0<sickdays<8	0<sickdays<11	0<sickdays<14	0<sickdays<8	0<sickdays<11	0<sickdays<14
VARIABLES	sick7	sick10	sick13	sickdays	sickdays	sickdays
female	0.0322***	0.0327***	0.0333***	0.0141	-0.0411	-0.0284
	[0.00217]	[0.00219]	[0.00220]	[0.0397]	[0.0451]	[0.0480]
age2529	0.00685**	0.00804***	0.00763**	-0.131**	-0.0615	-0.103
	[0.00299]	[0.00306]	[0.00307]	[0.0656]	[0.0766]	[0.0800]
age3034	0.00148	0.00198	0.00219	-0.0272	0.00289	0.0294
	[0.00272]	[0.00278]	[0.00280]	[0.0621]	[0.0720]	[0.0768]
age4044	-0.00531**	-0.00540**	-0.00510*	-0.0715	-0.0538	-0.00221
	[0.00265]	[0.00271]	[0.00272]	[0.0624]	[0.0725]	[0.0778]
age4549	-0.0141***	-0.0139***	-0.0138***	-0.0093	0.0618	0.0954
	[0.00276]	[0.00282]	[0.00284]	[0.0682]	[0.0799]	[0.0849]
age5054	-0.0189***	-0.0200***	-0.0202***	-0.0159	-0.0395	-0.0406
	[0.00276]	[0.00281]	[0.00283]	[0.0691]	[0.0793]	[0.0839]
age5559	-0.0256***	-0.0266***	-0.0267***	-0.0266	-0.0132	0.0231
	[0.00286]	[0.00292]	[0.00293]	[0.0737]	[0.0858]	[0.0918]
age6064	-0.0365***	-0.0368***	-0.0371***	-0.0216	0.135	0.16
	[0.00314]	[0.00321]	[0.00323]	[0.0940]	[0.114]	[0.121]
ptjob	0.00609**	0.00636**	0.00631**	0.0398	0.0378	0.0168
	[0.00259]	[0.00263]	[0.00264]	[0.0543]	[0.0628]	[0.0665]
immigrant	-0.0174***	-0.0176***	-0.0180***	-0.0426	0.022	-0.00533
	[0.00238]	[0.00243]	[0.00244]	[0.0629]	[0.0736]	[0.0773]
minority	-0.0105***	-0.0107***	-0.0109***	0.0201	0.0465	0.0388
	[0.00272]	[0.00277]	[0.00279]	[0.0693]	[0.0801]	[0.0851]
somehs	-0.0220***	-0.0226***	-0.0230***	0.297***	0.308***	0.272***
	[0.00230]	[0.00234]	[0.00235]	[0.0705]	[0.0810]	[0.0842]
hsgrad	-0.0153***	-0.0158***	-0.0163***	0.0345	0.0457	0.0108
	[0.00196]	[0.00199]	[0.00200]	[0.0510]	[0.0593]	[0.0619]
someps	0.00886***	0.00882***	0.00854***	-0.0436	-0.0737	-0.104
	[0.00317]	[0.00321]	[0.00322]	[0.0664]	[0.0753]	[0.0789]
chronic	0.0877***	0.0936***	0.0956***	0.195***	0.294***	0.331***
	[0.00310]	[0.00317]	[0.00319]	[0.0509]	[0.0599]	[0.0641]
child	-0.00242	-0.0033	-0.00375*	-0.0806	-0.131**	-0.164***
	[0.00209]	[0.00213]	[0.00214]	[0.0496]	[0.0567]	[0.0598]
single	0.00581**	0.00666***	0.00681***	-0.00947	0.0419	0.0471
	[0.00246]	[0.00251]	[0.00253]	[0.0554]	[0.0647]	[0.0689]
singleparent	0.00428	0.00451	0.00489	0.0421	0.0567	0.092
	[0.00451]	[0.00460]	[0.00463]	[0.0974]	[0.113]	[0.123]
inc014	0.00880***	0.00888***	0.00985***	0.206***	0.159**	0.220**
	[0.00316]	[0.00320]	[0.00324]	[0.0712]	[0.0800]	[0.0874]
inc1529	-6.58E-06	-0.000493	5.89E-05	-0.00456	-0.0413	0.0259
	[0.00212]	[0.00216]	[0.00217]	[0.0493]	[0.0563]	[0.0615]
inc5079	0.00163	0.00112	0.000712	-0.0555	-0.108*	-0.139**
	[0.00210]	[0.00213]	[0.00214]	[0.0491]	[0.0559]	[0.0586]
inc80plus	-0.00391	-0.00475	-0.00496	-0.113	-0.162*	-0.160*
	[0.00301]	[0.00305]	[0.00306]	[0.0741]	[0.0838]	[0.0889]
propinc	0.00378	0.00323	0.00317	0.202**	0.138	0.102
	[0.00404]	[0.00411]	[0.00413]	[0.0936]	[0.107]	[0.114]
Province fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	96705	97450	97723	12588	13333	13606

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 8: Time Senitvity Analysis - Females

	Probit marginal effects			Zero-truncated negative binomial marginal effects		
	0<sickdays<8	0<sickdays<11	0<sickdays<14	0<sickdays<8	0<sickdays<11	0<sickdays<14
VARIABLES	sick7	sick10	sick13	sickdays	sickdays	sickdays
age2529	0.0156***	0.0164***	0.0153***	-0.105	-0.104	-0.191*
	[0.00498]	[0.00509]	[0.00513]	[0.0846]	[0.100]	[0.107]
age3034	0.00108	0.000288	0.000483	0.0322	-0.0252	-0.00324
	[0.00450]	[0.00460]	[0.00465]	[0.0815]	[0.0957]	[0.105]
age4044	-0.00478	-0.0056	-0.00515	-0.0336	-0.0719	-0.00288
	[0.00453]	[0.00464]	[0.00468]	[0.0830]	[0.0982]	[0.109]
age4549	-0.0116**	-0.0124**	-0.0126***	-0.0239	-0.0269	-0.00946
	[0.00472]	[0.00484]	[0.00488]	[0.0896]	[0.106]	[0.115]
age5054	-0.0184***	-0.0211***	-0.0218***	-0.0533	-0.169	-0.183
	[0.00470]	[0.00480]	[0.00485]	[0.0891]	[0.104]	[0.113]
age5559	-0.0289***	-0.0314***	-0.0317***	-0.0947	-0.169	-0.125
	[0.00485]	[0.00497]	[0.00502]	[0.0957]	[0.113]	[0.124]
age6064	-0.0413***	-0.0442***	-0.0445***	0.0202	-0.0524	0.000996
	[0.00547]	[0.00560]	[0.00566]	[0.126]	[0.145]	[0.161]
ptjob	0.0052	0.00607*	0.00633*	0.0229	0.0629	0.0601
	[0.00354]	[0.00364]	[0.00367]	[0.0621]	[0.0748]	[0.0819]
immigrant	-0.0210***	-0.0210***	-0.0221***	-0.0419	0.0515	-0.0293
	[0.00397]	[0.00408]	[0.00411]	[0.0840]	[0.101]	[0.106]
minority	-0.0144***	-0.0146***	-0.0150***	0.0296	0.0748	0.0607
	[0.00446]	[0.00458]	[0.00462]	[0.0900]	[0.107]	[0.117]
somehs	-0.0242***	-0.0262***	-0.0271***	0.386***	0.285**	0.231*
	[0.00410]	[0.00420]	[0.00423]	[0.1000]	[0.112]	[0.120]
hsgrad	-0.0198***	-0.0208***	-0.0212***	0.00655	0.000893	-0.00897
	[0.00323]	[0.00331]	[0.00334]	[0.0658]	[0.0786]	[0.0861]
someps	0.0134**	0.0131**	0.0126**	0.0112	-0.0648	-0.118
	[0.00528]	[0.00538]	[0.00540]	[0.0844]	[0.0961]	[0.103]
chronic	0.116***	0.124***	0.126***	0.230***	0.320***	0.319***
	[0.00491]	[0.00499]	[0.00501]	[0.0704]	[0.0853]	[0.0937]
child	-0.00118	-0.0024	-0.00299	-0.125*	-0.181**	-0.214**
	[0.00360]	[0.00368]	[0.00371]	[0.0660]	[0.0779]	[0.0847]
single	0.0101**	0.0103**	0.00994**	0.0217	0.0124	-0.0232
	[0.00445]	[0.00456]	[0.00459]	[0.0777]	[0.0924]	[0.100]
singleparent	0.00292	0.00387	0.00459	0.0422	0.0971	0.147
	[0.00632]	[0.00651]	[0.00661]	[0.112]	[0.136]	[0.153]
inc014	0.0113**	0.0114**	0.0124**	0.158*	0.13	0.176
	[0.00470]	[0.00480]	[0.00486]	[0.0852]	[0.0985]	[0.110]
inc1529	-0.00319	-0.00398	-0.00372	0.000192	-0.0386	0.00459
	[0.00323]	[0.00330]	[0.00333]	[0.0600]	[0.0704]	[0.0785]
inc5079	0.00288	0.00105	0.000576	-0.0229	-0.146*	-0.183**
	[0.00380]	[0.00387]	[0.00390]	[0.0686]	[0.0793]	[0.0858]
inc80plus	0.00213	0.000216	-0.00116	-0.287**	-0.402***	-0.484***
	[0.00683]	[0.00694]	[0.00697]	[0.119]	[0.136]	[0.145]
propinc	0.00872	0.00989	0.0108	0.173	0.218	0.246
	[0.00692]	[0.00710]	[0.00711]	[0.125]	[0.148]	[0.162]
Province fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	47634	48065	48242	7527	7958	8135

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 9: Time Senitvity Analysis - Males

	Probit marginal effects			Zero-truncated negative binomial marginal effects		
	0<sickdays<8	0<sickdays<11	0<sickdays<14	0<sickdays<8	0<sickdays<11	0<sickdays<14
VARIABLES	sick7	sick10	sick13	sickdays	sickdays	sickdays
age2529	-0.000347	0.00132	0.00154	-0.172	-0.0311	0.00672
	[0.00473]	[0.00478]	[0.00480]	[0.107]	[0.119]	[0.124]
age3034	0.00268	0.0045	0.00477	-0.111	0.0287	0.0556
	[0.00433]	[0.00439]	[0.00440]	[0.0969]	[0.107]	[0.112]
age4044	-0.00739*	-0.00671	-0.00645	-0.131	-0.0341	-0.00964
	[0.00409]	[0.00412]	[0.00414]	[0.0957]	[0.105]	[0.109]
age4549	-0.0206***	-0.0193***	-0.0188***	0.00522	0.17	0.223*
	[0.00429]	[0.00433]	[0.00435]	[0.107]	[0.120]	[0.126]
age5054	-0.0254***	-0.0249***	-0.0248***	0.0287	0.132	0.152
	[0.00433]	[0.00436]	[0.00437]	[0.112]	[0.123]	[0.128]
age5559	-0.0308***	-0.0304***	-0.0303***	0.0534	0.186	0.211
	[0.00450]	[0.00452]	[0.00454]	[0.117]	[0.132]	[0.138]
age6064	-0.0444***	-0.0420***	-0.0423***	-0.105	0.359**	0.350*
	[0.00480]	[0.00486]	[0.00487]	[0.144]	[0.181]	[0.185]
ptjob	0.0172***	0.0157**	0.0148**	0.168	-0.00784	-0.0994
	[0.00630]	[0.00621]	[0.00618]	[0.118]	[0.120]	[0.119]
immigrant	-0.0193***	-0.0196***	-0.0195***	-0.0444	-0.015	0.0251
	[0.00383]	[0.00382]	[0.00383]	[0.0957]	[0.105]	[0.113]
minority	-0.0105**	-0.0107**	-0.0108**	-0.00341	-0.00467	-0.0112
	[0.00444]	[0.00444]	[0.00444]	[0.110]	[0.118]	[0.122]
somehs	-0.0261***	-0.0254***	-0.0255***	0.220**	0.323***	0.309***
	[0.00343]	[0.00343]	[0.00343]	[0.103]	[0.114]	[0.116]
hsgrad	-0.0151***	-0.0151***	-0.0157***	0.0924	0.123	0.0551
	[0.00315]	[0.00316]	[0.00316]	[0.0826]	[0.0896]	[0.0895]
someps	0.00688	0.00704	0.00688	-0.127	-0.0801	-0.0885
	[0.00499]	[0.00499]	[0.00499]	[0.108]	[0.120]	[0.123]
chronic	0.0820***	0.0868***	0.0889***	0.147**	0.235***	0.308***
	[0.00427]	[0.00433]	[0.00436]	[0.0738]	[0.0806]	[0.0844]
child	-0.00336	-0.00394	-0.00429	-0.0227	-0.0668	-0.0957
	[0.00327]	[0.00327]	[0.00328]	[0.0775]	[0.0822]	[0.0845]
single	0.00327	0.00512	0.00561	-0.0253	0.1	0.132
	[0.00365]	[0.00368]	[0.00369]	[0.0828]	[0.0907]	[0.0948]
singleparent	0.00526	0.000572	0.000363	-0.00361	-0.36	-0.325
	[0.0113]	[0.0110]	[0.0110]	[0.255]	[0.223]	[0.246]
inc014	0.00794	0.00924	0.0111*	0.361**	0.352**	0.492***
	[0.00641]	[0.00642]	[0.00647]	[0.141]	[0.154]	[0.170]
inc1529	0.00596	0.00557	0.00678*	-0.0326	-0.0491	0.0811
	[0.00395]	[0.00394]	[0.00397]	[0.0897]	[0.0960]	[0.104]
inc5079	0.00237	0.00281	0.00248	-0.0824	-0.0613	-0.0755
	[0.00312]	[0.00313]	[0.00313]	[0.0718]	[0.0770]	[0.0788]
inc80plus	-0.00271	-0.00263	-0.00221	-0.0381	-0.0165	0.0394
	[0.00416]	[0.00417]	[0.00418]	[0.101]	[0.107]	[0.113]
propinc	-0.00888	-0.0110*	-0.0120*	0.151	0.011	-0.0655
	[0.00668]	[0.00667]	[0.00668]	[0.154]	[0.163]	[0.169]
Province fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49071	49385	49481	5061	5375	5471

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

REFERENCES

Allebeck, P., & Masterkassa, A. (2004). "Chapter 5. Risk factors for sick leave – general studies." *Scandinavian Journal of Public Health*, 32(Suppl 63), 49-108.

Cameron, C. A., & Trivedi, P. K. (2009) *Microeconometrics: Methods and Applications*. (New York, NY: Cambridge University Press).

Canadian Centre for Occupational Health and Safety. (2008). *Common Cold*. Ottawa, ON. Retrieved from http://www.ccohs.ca/oshanswers/diseases/common_cold.html

Chevalier, A., & Viitanen, T. K. (2002). "The causality between female labour force participation and the availability of childcare." *Applied Economics Letters*, 9(14), 915-918.

Grootendorst, P. (1995). "A Comparison of Alternative Models of Prescription Drug Utilization." *Health Economics*, 4, 183-198.

Jones, A. M. (2000) "Health Econometrics." *Handbook of Health Economics*, A.J.Culyer & J.P. Newhouse (eds.), (Amsterdam: Elsevier).

Marshall, K. (2006). "On sick leave." *Perspectives on Labour and Income*. 18(2), 29-37. (Statistics Canada, Catalogue no. 75-001- XIE).

Public Health Agency of Canada. (2010). *Influenza*. Ottawa, ON. Retrieved from http://www.phac-aspc.gc.ca/influenza/index-eng.php

Statistics Canada. (2005a). "Canadian Community Health Survey 2003 User Guide for Public Use Microdata File." Ottawa, ON: Health Statistics Division, Catalogue no. 82M0013GPE.

Statistics Canada. (2005b). "Canadian Community Health Survey (CCHS) Cycle 2.1 Public Use Microdata File (PUMF) Derived and Grouped Variable Specifications." Ottawa, ON: Health Statistics Division.

Statistics Canada. (2005c). "Canadian Community Health Survey (CCHS) Questionnaire for Cycle 2.1 January 2003-November 2003." Ottawa, ON: Health Statistics Division.

Statistics Canada. (2005d). "Canadian Community Health Survey Cycle 3.1 Final Questionnaire." Ottawa, ON: Health Statistics Division.

Statistics Canada. (2005e). "CCHS Cycle 2.1: Data Dictionary Public Use Microdata File." Ottawa, ON: Health Statistics Division.

Statistics Canada. (2006a). "Canadian Community Health Survey (CCHS) Cycle 3.1 (2005) Public Use Microdata File (PUMF) User Guide." Ottawa, ON: Health Statistics Division.

Statistics Canada. (2006b). "Canadian Community Health Survey (CCHS) Cycle 3.1 (2005) Public Use Micro Data File (PUMF) Integrated Derived Variable (DV) and Grouped Variable Specifications." Ottawa, ON: Health Statistics Division.

Statistics Canada. (2006c). "CCHS Cycle 3.1: Data Dictionary Public Use Microdata File." Ottawa, ON: Health Statistics Division.

Statistics Canada. No date. *Days lost per work cause, by province (Illness or disability)* (table). Last updated March 24, 2009. Retrieved from http://www40.statcan.ca/l01/cst01/labor60b-eng.htm

Thomas, S., & Wannell, B. (2009). "Combining cycles of the Canadian Community Health Survey." *Health Reports*, 20(1). (Statistics Canada, Catalogue no. 82-003-XPE).

Van Poppel, M. N. M., de Vet, H. C. W., Koes, B. W., Smid, T., & Bouter, L. M. (2002). "Measuring sick leave: a comparison of self-reported data on sick leave and data from company records." *Occupational Medicine*, 52(8), 485-490.

Vistnes, J. P. (1997). "Differences in Days Lost from Work Due to Illness." *Industrial and Labor Relations Review*, 50(2), 304-323.