School of Business and Economics

Using graded sick leaves to affect leave length

An empirical analysis on the effects of graded sick leaves on leave length in Troms, Norway

—

Vidar Kristoffersen

Master's thesis in Economics – May 2019

Table of Contents

List of Tables

Foreword

Delivery of this master’s thesis marks the end of a five-year journey through higher education at University of Tromsø’s School of Business and Economics. The road has been long, sometimes arduous and for the most part enjoyable.

I would like to extend my gratitude to my supervisor, Mikko Moilanen, for being exceptionally helpful when we began planning for this thesis last semester. Thanks to his foresight and support I was able to bounce back from an unfortunate roadblock that necessitated a change in thesis topic and complete this thesis in the normal amount of time. I would also like to extend a thanks to Øystein Myrland who helped a lot with the technical aspect of cleaning up the data and developing the model in R.

Lastly, I would like to thank my parents, without their support these last five years would have been significantly harder.

Abstract

Graded sick leaves might be used to extract otherwise untapped capacity for productivity in sick or injured workers, but what are the secondary effects of such a policy? This thesis aims to answer the question of how the use of graded sick leaves might affect the length of the sick leave using comprehensive sick leave data from Troms county in Norway from 2010 through 2017. A parametric Weibull survival model is used in an attempt to isolate the choice of grading or no grading and analyse the effects. The analysis results show that being granted a graded sick leave tends to increase the actual length of medium- and long-term sick leaves. For effective sick leave length (defined as the product of sick leave days and the average grading of the leave) being granted a graded sick leave is a positive benefit when we exclude medium-length sick leaves of 8 weeks or shorter, reducing the effective sick leave length and thereby contributing to an increase in the overall productivity. Due consideration must be given before extending these results beyond the analysis done in the thesis because of a lack of diagnosis data.

Keywords: graded sick leaves, sick leave length, effective sick leave length, survival analysis

1 Introduction

When we look at how our labour market functions and the design of the various mechanics we have implemented into our labour market to deal with situations beyond the ordinary, we begin with a default starting point where the workers are working at their full expected capacity at their job, they clock in when they are expected to begin working and clock out when the pre-determined amount of work has been done or the pre-determined amount of time has passed by. The employers pay their workers their pre-determined and hopefully fair wage for their work. A worker might quit and change their employer or change their vocation altogether, and the employer replaces that worker with a new one that, in this simplified model, works at the same capacity as the old one.

An unexpected event may occur, an accident or a sudden sickness, that reduces a worker's capacity for work. There are solutions to this, the simplest one simply copying the strategy for when an employee quits: take the worker in question out of his workplace and replace him with a new worker whose capacity for work is full, as expected of any worker. This solution has two problems but solves only one. First, the employer experiences a vacuum left by the sick employee, this vacuum is easily filled by the replacement worker. The second problem is that, assuming the workers depend on their work to provide the income they need to sustain themselves, the sick worker is left without any income. With no intervention the sick worker will likely die of either hunger, exposure or other dangers that present themselves when left with no way to provide for themselves.

A secondary solution that solves this second problem can easily be developed, and it's something we see in many countries today: the government intervenes in some capacity to ensure that the sick worker is still able to cover his basic necessities for life. This intervention can take many forms: there might be a law that mandates that an employer keeps a sick employee on the payroll rather than outright fire him or the government might themselves pay a portion of the sick worker's lost wage, or some combination of these two and other solutions. Either way the worker is provided for and won't die due to his bad luck through this sickness benefits program.

A more elegant solution, one that solves both problems presented by the unexpected sickness, presents itself if we are willing to look at a case-by-case basis and have someone capable of correct evaluation of each individual worker's unexpected sickness or injury: graded sick leaves. If we assume that the sickness doesn't completely eradicate the worker's capacity for work, but instead reduces it by some percentage; a person whose job only consists of sitting at a desk and doing work there might only lose 10% - 20% of his capacity for work if he breaks a leg, while a hiking tour guide would lose 100% of his capacity for work with the same injury. A person inflicted with chronic fatigue syndrome (CFS) might have their capacity for any type of work heavily restricted, but they might still be capable of working a couple of hours a day before they find themselves exhausted – meaning their capacity for work is reduced by 75%. It stands to reason that, if the workers can provide some amount of work at the quality that is expected of them, they should be allowed to provide that work and receive compensation that is adjusted to what they provide. Assuming that their capacity for work still leaves them with enough income to ensure that their basic necessities are covered, the government might not even need to step in. However, in the case of the person with CFS and similar cases, it is likely that 25% of their standard income is not enough to cover the necessities, so it might be required to mix this solution with the solution of sick benefits discussed earlier.

Using graded sick leaves seems like an ideal solution to an awkward problem that can happen with anyone and with a large associated loss of productivity and efficiency. Not only does it allow for the extraction of the amount of productivity an injured or sick worker is capable of providing, you potentially avoid the added cost of providing benefits to the worker while he heals – or at least part of it, if you end up with a solution where the worker is given a percentage of his wage larger than his capacity for work.

So far, it has been assumed that the worker's capabilities are accurately assessed, while that may not necessarily be the case. Two problems arise in doing this assessment, no matter who does it. The first is accurately determining the extent of how much the disease or injury affect the worker's capabilities, and accurately determining which capabilities are relevant to the tasks that the worker is responsible for in his work environment. This was touched upon earlier; a broken leg presents significantly more of a problem for a hiking tour guide that relies on his feet to cover rough terrain than it does for a desk worker who doesn't need to move around to perform his work duties. In this particular example, how the injury affects the worker's capabilities should be obvious to most people and does not require special consideration when determining who performs the assessment, but more complicated situations can quickly arise – for example, how much does an ongoing cancer treatment affect working capabilities? How much does a chronic, worsening muscle dystrophy diagnosis like ALS affect the person's capabilities depending on how far along the diagnosis is? These questions make it obvious that a physician should have a major part of performing this assessment. The second problem, which capabilities are important for the execution of work duties, are, if not more obvious to most people, at least easier and quicker to learn than an education in medicine. Hence, for the most accurate assessment, it makes sense for a physician to do it in conjunction with the worker/workplace for work-specific information. While the assessment will likely never be 100% accurate, it doesn't have to be as long as the physician errs on the side of caution, as it will probably be better for the worker's overall prognosis if he works slightly less than what he is capable of rather than slightly more.

An interesting question that occurs in relation to this program is what the unintended or potentially intended secondary consequences are. If we define the primary purpose of using graded sick leaves as extracting productivity from a worker with a reduced capacity to work, which it accomplishes as discussed previously, we could look at the other potential effects of graded sick leaves as secondary purposes. One of the more interesting side-effects is how they impact the length of the sick leave itself, as one might expect there to be some noticeable effect on how long a worker retains his disease or injury and until his working capacity is fully restored. Intuitively there are several potential consequences that could make the overall effect lead to either an increased or a decreased sick leave length. If the activity slows down the healing, you would expect a graded sick leave to be longer than an ungraded one – although, from a societal point of view, that might still be a net positive if the overall productivity remains higher, that is to say 10 days of 55% work participation is preferable to 5 days of 0% work participation from a point of view of pure productivity. On the other side, if being active and included promotes healing of the injury or sickness then using graded sick leaves will unequivocally be positive all around. Likewise, one could construct a hypothetical where a worker with a long-term disability that is not given graded sick leaves could fall out of the workforce altogether because of how long he is taken out of work for, while an identical person on graded sick leaves remains active in the workplace and social with his co-workers, making it significantly less likely that he falls out of the workforce.

In summary, graded sick leaves are an attractive option to the binary 0%/100% sick leave system, for potentially more reasons than just the extraction of available productivity. In Norway, you can give yourself a temporary, limited sick leave (self-certified sick leave) in some limited circumstances ("Egenmelding," 2019). You can only do so for the first 16 calendar days of your absence from work in a given year, and you must have worked for your employer for at least two months. Furthermore, if your self-certified sick leave lasts for more than 3 consecutive days your employer can require you to submit a physician-certified sick leave, presumably to give the employer an easy way to verify the system is not being abused. Critically, a self-certified sick leave cannot be used as a graded sick leave. Overall, the system of self-certification of sick leaves is likely included as a way to avoid needlessly encumbering the various doctor's offices around the country for simple, common illnesses like the common cold, that most workers can typically expect to experience once a year or on that order of frequency, while making a few concessions to minimize the potential for abuse by the employees. Ultimately, for the purposes of this thesis doctor-certified sick leaves are of significantly more interest. It is however worth noting that due to this system, the injuries/sicknesses for which a worker is inclined to use the self-certified sick leave system is very unlikely to be used with the graded sick leave system – methodologically this impacts how a study of graded sick leaves wants to filter their dataset for accurate results, discussed more in depth in the data section.

Sick leaves where a doctor has examined the worker and issued a corresponding sick leave have more options compared to the self-certified ones. This thesis will focus on the option of grading the sick leave, that is to allow the worker to take some percentage of time off work to recover from his sickness or injury. ("Gradert sykemelding," n.d.) In Norway, the examining physician is the person who decides to what degree the sick leave should be graded with supplemental information regarding the specifics of the workplace and furthermore, it is the preferred type of sick leave when a doctor is the one issuing it, assuming it is an available and viable option for the worker and the workplace. The lowest grade a sick leave can get is 20%, with the highest naturally being 100% i.e. a non-graded sick leave. The worker is compensated for the amount of work he does under his graded sick leave and receives sickness benefits to cover the remaining portion of his work that he's granted leave from. The worker and the employer work together to figure out the best way to accommodate for and execute the completion of the employee's work responsibilities, whether that be through giving the worker less hours and possibly responsibilities at work or through giving the worker more time to accomplish fewer responsibilities (for example, a person with a sick leave graded at 50% with a standard work day of 8 hours can either work at 100% capacity for 4 hours then take the remaining 4 hours off or work at 50% capacity for the full 8 hours) or a combination of the two, depending on the specifics of the worker and the nature of his sickness or injury.

In general, it appears Norway has higher rates of sick leaves than other countries when compared with similar criteria and definitions of sick leaves (Berge, 2012). The consequence of this is that Norway would benefit more when compared to other countries when it comes to designing, applying and evaluating the effects of various sick leave programs (such as graded sick leaves) that improve upon the basic binary sick leave system. Looking at the actual numbers, below is a table with data from SSB showing how big the portion of the work force granted a doctor certified sick leave was during the four quarters of 2018 for all occupations ("12452: Doctor certified sickness absence for employees (per cent), by occupation, quarter, sex and contents," n.d.).

Table 1: doctor certified sickness absence rates

	Males and females	**Males**	**Females**
2018 Q1	5.5%	4.2%	7.2%
2018 Q2	4.8%	3.6%	6.2%
2018 Q3	4.2%	3.3%	5.3%
2018 Q4	5.1%	3.8%	6.7%

As is evident from a surface-level inspection of the numbers, women in general seem to have much higher rates of sickness absences from the workforce compared to men, but both sexes have a sizable population that at any given time is granted a doctor certified sick leave, as such the benefits of utilizing a system like the graded sick leave system would be large on a societal level. A simple numerical example follows: if all the sick leaves were graded at 100%, the country would lose an average of 4.9% of the total worker productivity throughout the year – stating it differently, the total worker productivity is 95.1%. If the people granted sick leaves have an average remaining working capacity of 30% even with the sickness or injury they are granted leave for, the country would have a total worker productivity of 96.57, in improvement of 1.47 percentage points that is entirely "free". The assumption of an average working capability of 30% could be way off as well, with naturally higher gains if the real number is higher. Furthermore, as this thesis will focus on the secondary effects on leave length rather than the actual productivity gains from grading sick leaves, there might be additional effects that further increase or potentially decrease the gains from using the graded sick leaves.

Formally, it makes sense to separate the potential effects of graded sick leave on leave length into several possibilities: first, it has been shown that, for people with disabilities and illnesses that are viable options for graded sick leave, work-related activity increases rates of recovery and rehabilitation, generally leads to better health outcomes for the individuals, minimizes the harmful physical, social and mental side-effects of long-term absence and reduces the risk of the individuals experiencing long-term incapacity in the workforce (Waddell & Burton, 2006). From an economics point of view, these are all positive effects that will contribute to a reduction in absence length and possibly a reduction in future absences as well.

An additional effect that could present itself is the risk inherent in returning to work too early, that is before the individual has had sufficient time for healing and rehabilitation of their cause of absence. Claims that returning to work too early leads to a higher risk of future sickness absences have been made, but have as of yet not been proven empirically (Bruusgaard & Claussen, 2010) (Brage, Kann, Kolstad, Nossen, & Thune, 2011). Concerns have also been raised that allowing ill and injured people graded sick leaves can lead to a lock-in effect, wherein a person who is currently on a graded sick leave could be tempted to reduce their own effort to heal and return to full working capacity (Brage et al., 2011), presumably because they prefer how graded sick leave works out in terms of their working hours and patterns with the accompanying compensation. Economically, these are effects that would contribute to a lengthening of the sick leave and an increase in future sickness absences, should they manifest.

This thesis aims to answer the following question: what effect, if any, is there on the length of any given work absence with a graded doctor certified sick leave when compared to a situation where the same absence was given a non-graded doctor certified sick leave. To answer the question, a dataset containing all doctor certified sick leaves in Norway from 2010 through 2017 supplied by the Norwegian Labour and Welfare Administration (NAV) will be used to develop a regression analysis that will attempt to isolate the effects of graded versus non-graded sick leaves. This dataset is obviously observational in nature, due to the ethical concerns as well as difficulty in using any type of experimental approach when attempting to answer this question.

The rest of the thesis is organized as follows: first a general literature review will give an overview of the research that has already been done on the subject, giving context to the thesis and see where it fits in the current research landscape. The next section details the dataset that has been supplied by NAV with some descriptive statistics, as well as describing how the dataset has been limited or otherwise modified in order to improve the accuracy and reliability of the regression analysis and subsequent results. Following the section on data, an overview of the theoretical framework behind the statistical techniques used to obtain the results presented in this thesis is given. Following this is a section on the empirical analysis including the procedure of the regression analysis as well as presenting and discussing the results. The thesis wraps up with concluding remarks.

2 Literature review

There has been done a surprising amount of research on the question of how graded sick leaves (also known as partial sick leaves) affect the length of the sick leave versus non-graded or full sick leaves. These studies all display varying levels of quality regarding the method they have chosen to analyse the data, how they have decided to modify the data prior to analysis as well as the dataset itself and their base assumptions and definitions. This literature review aims to give a brief overview of key papers in this field of economic research, assessing their strengths and weaknesses, highlighting key differences in their approach to the problem and ultimately compare their results. The literature review is limited to studies from Nordic countries, given their general similarity in the structure of their social programs and societal norms.

2.1 Observational versus experimental data

The vast majority of the studies done on the subject have used observational data, that is data that is naturally generated through the execution of pre-existing, independent policies or otherwise naturally occurring situations. In essence, the people who wish to investigate how graded sick leaves affect leave length have to take what data they can get, data that has been generated when the governing body of the country they're looking at applies graded sick leaves to their citizens in their attempts at improving citizen's outcomes. The flip-side of observational data is experimental data, that is an arranged data generation event or series of events where all (or as many as possible) other variables are held constant. The variable the researchers are interested in, in the case of this thesis the use of graded sick leaves, are altered and the resulting dependent variable (in this case, the length of the sick leave) is observed. This controlled environment allows the researchers to draw causal interferences from their data (Croson & Gächter, 2010).

Performing economic experiments tend to carry practical as well as ethical problems. In the case of sick leaves, if one were to attempt a controlled experiment, you would have to recruit a number of people with injuries or sicknesses that leave them some capacity for work, trying to control for all the other various factors that might impact the length of the sick leave (occupation, gender, work environment, recovery environment and so forth) then randomly assign them either graded or non-graded sick leaves and observe the length of the sick leave. This would necessitate not only removing the participants from the pre-existing sick leave programs in the country of the experiment, programs which generally aim to give the citizens who need it the best outcome possible with the current knowledge available. It would also risk leaving a participant in a worse state than they otherwise would have been in, practically speaking a participant may find themselves experiencing a longer sick leave due to participating in the experiment.

The more common option of using observational data carries problems related to the reliability of the analysis. Typically, you encounter issues with the counterfactual, so while you may know what happens if individual A receives graded sick leave, you do not know what happens if individual A (or an individual practically identical in all relevant aspects) does not receive graded sick leave. These problems can be alleviated to some extent when the dataset is sufficiently large in individuals and observations, so that there might by random chance be individuals who are very similar and only differ in whether they received a graded sick leave or not. This does not mean that an exceptionally large observational dataset is equivalent to an experimental data set however, as often the treatment (graded sick leaves) will have selection bias. A person in an occupation that can't work remotely who is granted a sick leave for a particularly infectious illness is unlikely to ever be considered for a graded sick leave, likewise someone who injured their foot in an occupation that relies on foot movement to a limited extent is unlikely to ever receive a full, non-graded sick leave. The random pairings described earlier won't occur in these situations no matter how many individuals or observations you have.

Furthermore, with observational data the researcher can encounter problems with unobserved heterogeneity among the individuals. The dataset that the researcher receives might contain most of the information regarding relevant variables, but some might be unavailable due to concerns and legislation surrounding for example privacy or just practical issues with collecting or measuring certain variables. As examples, the observational data for sick leaves may lack information regarding the participant's diagnosis or earlier health history given that health information is often very heavily protected in legislation. Or it may lack information regarding the environment the individuals find themselves in because of the practical difficulty in assessing and quantifying work and home environments even though it may impact how long a sick leave lasts for and thus be an important variable in the analysis.

2.2 Studies

2.2.1 Evidence from Finland

(Viikari-Juntura et al., 2012) The only study included in this literature review that used experimental data is a Finnish study. They recruited individuals with musculoskeletal disorders aged 18-60 that were unable to perform their work duties at 100% capacity. The researchers randomly assigned the individuals to either a full sick leave or a partial sick leave. The partial sick leaves consisted of 70% of individuals working about half of daily working time, the remaining 30% of individuals had difficulties in doing so and instead worked somewhat shorter hours 3-4 days a week. Furthermore, if necessary the work tasks they performed were modified to not risk exacerbation of symptoms related to activity. In total there were 62 participants in the experiment, with a 50% split between the treatment group and the control group.

To analyse the data, the researchers used a Cox proportional hazard model to estimate the hazard ratios for returning to work, that is how a variable changes the length of time until the event occurs (in this case, returning to work). The study found that the treatment group who received partial/graded sick leaves experienced a shorter sick leave duration when compared to the control group who was granted a full sick leave, both in the short term and the long term. Given that the study limited itself to participants who suffered from musculoskeletal disorders which likely has predictable patterns of symptoms and predictable effects on working capacity, one should be careful with extending this result to larger groups of injuries, sicknesses or disorders.

In terms of quality, a systematic mapping review done by Meneses-Eschavez, Baiju and Berg evaluates the experimental study to be of moderate quality, noting that the experiment failed to obfuscate control/treatment group assignment for the participants and the assessors, in addition to being susceptible to selective reporting of the outcomes (Meneses-Echavez, Baiju, & Berg, 2018). Furthermore, almost all the participants in the experiment were female which means the results fail to include any differences that might exist between how men and women react to graded sick leaves. Given that there are such large differences between men and women in the general sickness absence numbers such a factor could be highly important to the research question.

2.2.2 Evidence from Sweden

(Andrén, 2014) The only Swedish study included in this literature review was written by Daniela Andrén in 2014. The paper's goal is to give an answer to the question of how graded sick leaves can affect the probability that employees diagnosed with mental disorders returns to work – mental disorders being defined as a set of clinically recognizable symptoms that together form a mental disorder diagnosis defined by the medical classification list ICD-10. To perform the analysis the study used a 2002 data sample from the National Agency of Social Insurance in Sweden, the Swedish equivalent of NAV in Norway. The total number of individuals in the sample were 627, 79 of which (12.6%) was granted graded sick leaves from the start while the remaining 548 (87.4%) started on full sick leaves, forming the control group.

To analyse the data, the study developed a discrete choice switching regression model which included an endogenous switch between part-time sick leave and full-time sick leave. They predict the probabilities of four different events occurring when participating in the part-time sick leave program: successful (full recovery of lost work capacity due to part-time sick leave, full-time would not have led to recovery), positive indifference (part-time sick leave had no effect, the individual would have recovered anyways), negative indifference (part-time sick leave had no effect, the individual would not have recovered anyways) and unsuccessful (failure to recover lost work capacity, full-time sick leave would have led to recovery).

The results of the study were that for most cases, the effects of part-time sick leave participation on length of the sick leave was slightly either negative or positive, depending on the exact circumstances. A notable result found was that, for people with long-term sick leaves of 60 days or more who started their sick leave on a full-time basis but transitioned to part-time sick leaves, the participation in part-time sick leaves gave a strong positive effect in reducing the length of time of the sick leave. The implication being that for people who are disconnected form the workplace for a long period of time, getting them an active connection to the workplace has a strong positive effect on recovery (Andrén, 2014, p. 9). Like Viikari-Juntura et al. (2012), the scope of the population sample was significantly restricted so generalizations from the results are difficult to make. In terms of quality, Meneses-Echavez et al., (2018) rates the study of moderate quality.

2.2.3 Evidence from Denmark

(Høgelund, Holm, & McIntosh, 2010) The only Danish study examines the effect of a national graded sick leave program on the probability of the participants returning to work. The data used in this study is comprised of individuals with long-term sick leaves of 8 weeks or longer. The total sample was 934 individuals, with 265 in the treatment group with graded sick leaves and 669 in the control group with full sick leaves. The data included information regarding an individual's number of visits to their general practitioner during the previous year as well as whether the sick leave was caused by a mental disorder or if it had other causes.

To analyse the data, the paper used a multivariate mixed-proportional-hazard-rate model – a survival model similar to the one used by Viikari-Juntura et al (2012). To identify the treatment effect, the paper uses a timing-of-event approach that hinges on the assumption that the participating employees are unable to anticipate when the program participation begins. The paper finds that participating in a graded sick leave program yields significant positive results for the participant in terms of the probability of returning to work, in addition to finding a non-significant effect of program participation without returning to work. Quality-wise, a potential point of weakness is the limiting of the sample population to individuals with sick leaves longer than 8 weeks, as this could potentially exclude a number people with disorders/injuries/illnesses with shorter absences that are viable for graded sick leaves that could leave different results (Nossen & Brage, 2013). Meneses-Eschavez et al. (2018) assesses the paper at moderate quality.

2.2.4 Evidence from Norway

(Markussen, Mykletun, & Røed, 2012) The first of two papers from Norway, this study aims to answer whether using a strategy that attempts to exploit the remaining capacity for work that a sick or injured worker might have could reduce the amount of absenteeism, as well as reduce dependency on social welfare programs and thus promote self-sufficiency among the workforce. The data they used in this study was received from Norwegian administrative registers and includes data from 2001 and onwards – the data regarding physician-certified sick leaves spanning the time period of 2001-2005. The dataset includes basic demographic data regarding individuals who were granted sick leaves, notably they were able to obtain diagnosis data for each sick spell, income data per individual and data on which doctor certified the sick leaves. The total number of individuals in the treatment group that received graded sick leaves were 77 655 whereas the number of individuals in the control group numbered 261 596. They excluded any sick leaves that lasted shorter than 8 weeks and defined a sick leave as graded if the grading happened before the 8-week mark in the sick leave.

To analyse the data, the paper wanted to account for differing propensities among doctors to prescribe graded sick leave rather than a full sick leave and included this in their regression – this is worth noting as no other study included in this literature review has done this. The analysis takes form of an instrumental variable approach with a linear regression framework. Their findings indicate that grading sick leaves gives large economic contributions in the form of shorter sick leave length, a reduction in reliance on social security and an increase in employment propensities for individuals. The note that their results show using graded sick leaves significantly raises the individual's likelihood of remaining employed after the sick leave.

Meneses-Eschavez et al. (2018) rates the study of moderate quality. A potential weakness is their exclusion of any sick leave shorter than 8 weeks in their analysis, for the same reasons as Høgelund, Holm & McIntosh (2010). Nossen & Brage, 2013 also criticize the study for the same reason, noting that excluding sick leaves shorter than 2 weeks should be sufficient for removing the most egregious cases where graded sick leaves are not viable options. Furthermore, the data is somewhat old by now and limited in length, covering four years.

(Atle Lie, 2014) The final report included here is a second Norwegian report from 2014. It aims to compare the effects of graded versus full sick leaves on the duration of the sick leave. The data used consists of a sample of 10% of every person in Norway who was registered with at least one employment record from 2002 to 2010. For analysis, the report used an extended Cox proportional hazard model, a survival model that looked at the survival time of the sick leave, where the "death" was defined as the end of the sick leave, either a return to work or other social security programs. The report found that there were no clear effects of graded sick leaves on leave length versus full sick leaves. People who were given a graded sick leave from the first day of the sick leave risked longer sick leaves than the people who were given a full sick leave, with less clear results for other groups of people.

Meneses-Eschavez et al. (2018) rates the study of moderate quality.

2.2.5 Final notes on literature

Overall, most studies seem to predict a positive, either strong or somewhat limited, effect on sick leave length when the individuals on sick leave are given graded rather than full sick leaves with the exception of Lie (2014) who predicts negative to no notable effect. Two of the papers used survival analysis techniques, the remaining three used other regression analysis techniques to analyse the data. The quality of each paper has been acceptable with a few points of interest, especially regarding generalization of the results. The datasets used has been comparable, with an interesting addition of physician propensity in Markussen, Mykletun og Røed (2012). Comparatively, the dataset used in the analysis for this thesis is more lacking, this will be expanded upon in the data section.

3 Data

This section will detail exactly what data is being used in the analysis with descriptive statistics as well as all the modifications, restrictions and other relevant actions done to the data for various reasons. Additionally, it will contain a description of what data points are missing for the observations included and attempt to explain how big an impact they might have and how important these missing data points might be.

The raw data received from NAV are split into three parts. Throughout these three parts, individuals have been de-identified, however they are assigned a unique identification ID that is identical between data sets. The first and most important set is the sick leave data; this dataset contains information per individual for every doctor-certified sick leave that has occurred in Norway registered with an ending date of January 1. 2010 or later, as well as a beginning date of December 31. 2017. Effectively this means the vast majority of observed sick leaves occurred from 2010 through 2017 with a few outliers on either end of the time period. The dataset contains information on the individual’s start and stop date for their sick leave, whether they are male or female, the general category of their occupation (see appendix section 1 for details), the sector they work in (either private, state or municipality), their age grouping (5 year intervals, see appendix section 2 for details) and lastly the average grading of their sick leave. This means that the data set does not contain any information as to how the grading differed throughout the course of a given sick leave, unless the new grading was accompanied by a new sick leave, which is often not the case. Nor does the data set contain any information as to the number of times the grading changed, so a 10-day sick leave observation with an average grading of 60% could have had a single grading of 60% throughout the 10 days, 5 days of 90% grading and 5 days of 30% grading or any number of other combinations.

The second and third datasets were supplementary and only used for filtering and narrowing down the main data set with sick leave observations. The second dataset consisted of data on each individual's history with receiving either work assessment allowance – meaning they had a significant reduction in work capacity but did not receive sickness leave benefits (possibly because they had used up what benefits were available to them) - or unemployment benefits in any given month. The third dataset contained information on whether any given individual included in the first dataset with sick leaves had employment in any given month. These datasets were merged together to form a single dataset with information on every individual included in the first, primary dataset. In total there were 10 144 498 observations of 2 181 606 unique individuals. While the original goal of the analysis was to use data for the entirety of Norway, the resulting dataset was so large any calculations took an inordinate amount of time and the dataset itself took a significant amount of hard drive space. For this reason, the dataset was limited to only include sick leaves from Troms county. Limited to only Troms county, the total number of sick leave observations were a much more manageable 367 963 sick leaves across 75 784 unique individuals.

In order to accurately assess how graded sick leaves affected the rate at which the workers recovered (by ending the sick leave and returning to full-time work), the dataset could only include people whose sick leave ended with a full recovery, resuming the normal employment they likely had prior to what caused the sick leave to occur. The second and third dataset were therefore used to remove any sick leave that didn't conclude with a successful return to normal work but instead concluded with unemployment benefits, meaning they lost their jobs or quit or otherwise ended up unemployed. Likewise, we also filtered out any sick leave that ended with the individuals receiving work assessment allowance directly afterwards, as that would indicate that the person did not recover from his sickness or injury. Including these two categories of observations would skew the results.

Furthermore, any sick leaves with an end date later than 31. December 2017 were removed to avoid problems with right censoring, since the data didn't include any outcomes after this point – the end dates specified in these observations were merely administrative and predicted end dates rather than actual end dates. Four additional variables were created for each sick leave observation: length of sick leave in days, amount of observed sick leaves per individual, length of time since conclusion of last sick leave in days and a dummy variable indicating whether the sick leave was graded or not (value 1 for average grading < 1.0, otherwise 0). Any sick leave with length of 0 days or less (indicating either an error or a formality entry in the data, no actual sick leave occurred) and more than 365 days (indicating a chronic illness that is unlikely to be affected by grading the sick leave, and also very likely to receive a graded sick leave to begin with) were removed. 84 observations that lacked an entry for their occupation category were also removed. Any observations where the amount of days between sick leaves were 0 or less were removed to remove a few overlapping sick leaves, indicating errors in how their sick leaves were reported

Lastly the first observed sick leave for each individual (where amount of days since last sick leave was NA) was removed to avoid issues with left censoring since the variable "days since last sick leave" is used as an independent variable in the analysis. It's worth pointing out that this modification results in removing all the people who only had a single sick leave over the 8-year period. The final dataset used for the analysis consists of 203 948 sick leave observations between 45 363 unique individuals, with 154 257 sick leaves being full-time, non-graded sick leaves and the remaining 49 691 sick leaves being graded.

Table 2: descriptive statistics

	Number of sick leaves	Proportion
Grading		
Full-time sick leave	154257	75,64 %
Graded sick leave	49691	24,36 %
Gender		
Female	129207	63,35 %
Male	74741	36,65 %
Occupation		
Military/unspecified	2762	1,35 %
Legislators/leaders	9681	4,75 %
Technicians/associate professionals	42897	21,03 %

Professionals	34280	16,81 %
Clerks	12189	5,98 %
Service/shop/market sales workers	61564	30,19 %
Agriculutral/fishery workers	1220	0,60 %
Craft/related trades workers	15150	7,43 %
Plant and machine operators/assemblers	11312	5,55 %
Elementary occupations	12893	6,32 %
Sector		
Private	90175	44,21 %
State	42786	20,98 %
Municipality	70987	34,81 %
Age group		
20-24	15521	7,61 %
25-29	22431	11,00 %
30-34	24003	11,77 %
35-39	25230	12,37 %
40-44	27744	13,60 %
45-49	26281	12,89 %
50-54	22721	11,14 %
55-59	21070	10,33 %
60-64	18947	9,29 %

Table 3: more descriptive statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Average grading	0.909	0.186	0	1	1	1
Length of sick leave	35.961	64.475	1	4	33	365
Number of sick leaves	10.282	7.861	2	5	13	75
Days since last sick leave	277.802	360.585	1	48	355	2,859

A central question worth asking if it the dataset might be lacking certain data points that contribute to unobserved heterogeneity among the individuals. A key point of data that most of the previous studies done on this subject did have was the data regarding the diagnosis of the individual on sick leave. Unfortunately, due to the strict rules surrounding privacy regarding sensitive information like health data for individuals (even if it's de-identified), gathering this information was not viable given the limited amount of time allotted for a master's thesis. The lack of diagnosis data hurts the analysis in two ways: first, it's very likely a large influencer on the length of the leave time, meaning we lose a significant explanatory variable for our analysis. In addition, Nossen & Brage (2013) makes a series of suggestions regarding how to pick and modify your data set to improve results when analysing graded sick leave effects. Two of these suggestions rely on diagnosis data: first, they argue it makes sense to exclude all sick leaves with an accompanying diagnosis with a median length of below a given threshold, because if the typical prognosis for a given diagnosis is rapid recovery it makes less sense for the attending doctor to assign the individual graded sick leave – empirics show that graded sick leaves are practically never assigned for sick leaves with a duration of 1-16 days (Brage et al., 2011). This particular issue can be somewhat remedied by simply excluding any sick leave shorter than 14 days, which is an alternative Nossen & Brage (2013) suggests – they prefer excluding any sick leave 14 days or shorter because extending the exclusion length risks leaving out many sick leaves that weren't graded but very well could have been (Nossen & Brage, 2013). This analysis will look at both a 2-week and an 8-week exclusion length for sick leave time.

The second suggestion by Nossen & Brage (2013) is to exclude any sick leave where the diagnosis is related to pregnancy. Pregnant women, they argue, have a distinct pattern of absence. Intuitively it makes sense – no amount of grading a sick leave will shorten the time of a successful pregnancy, so their inclusion will likely skew the data towards no effect. Furthermore, pregnancy-related diagnoses are often given the largest percentage of graded sick leaves – 39% of their sick leaves are graded (Brage et al., 2011). This weakness in the data is much harder to alleviate by proxy without the actual diagnosis data short of removing every woman from the dataset – this was considered but ultimately not done for this analysis because of the observed differences in sick leave usage by men and women.

4 Method

This analysis will use a parametric survival model called Weibull where the dependant variable is the amount of time until a given event occurs, in this case the event is the conclusion of the sick leave and a return to normal work. The more widely used Cox proportional hazard model (which is a semi-parametric survival model) was first considered, however a diagnostic Schoenfeld residuals test returned significant results, suggesting that the model didn't fit the data very well (Moore, 2016, Chapter 7). A Weibull model was instead chosen. Furthermore, the data was pre-processed with a matching procedure called MatchIt, detailed below – the matching procedure did not impact the analysis results, but it remains included as a measure of the robustness of the data.

4.1 MatchIt

The statistical underpinnings of the MatchIt matching procedure is detailed by Ho, Imai, King, & Stuart (2011) and this entire section is based on their paper. A significantly more in-depth detailing can be found in Ho, Imai, King, & Stuart (2007). MatchIt is designed for use in analyses where the researcher is looking for causal interferences with a dichotomous treatment variable (in our case, whether the sick leave was graded or not) and a number of control variables. n_1 are the number of treated individuals, n_0 the number of controlled (untreated) individuals. The control variables are denoted by a vector x_i for individual i. When the individual i is assigned control, $t_i = 1$ and $t_i = 0$ when assigned treatment. The outcome is denoted by y_i, with $y_i(0)$ being the outcome of individual i under control and $y_i(1)$ the outcome of the same individual under treatment. Naturally, both outcomes cannot be jointly observed. Lastly, there are fixed vectors of pre-treatment, exogenous confounders denoted by X_i which are measured.

The point of the matching procedure is to match data so that any relationship between X_i and t_i is eliminated or at least minimized – assuming that is not already the case, in which case any analysis would be reduced to simply observing the differences in outcome y_i with differing treatment t_i and holding the control variables x_i constant. The MatchIt model attempts to do this without introducing bias and without introducing too much inefficiency. It removes, duplicates and selects specific observations through a rule that is only a function of the treatment t_i and the exogenous confounders X_i without depending on the outcome Y_i. The model makes use of a propensity score, calculated by the probability of being assigned treatment given the control variables. The model keeps all treated individuals and instead opts to weigh the control units, thereby not affecting the quantity of treated individuals who are of interest to the analysis.

While MatchIt offers several different matching techniques, the one used in this analysis is a subclassification procedure. While exact matching would have been ideal because it matches each treated individual to every control individual it was unfeasible given the large range of values the covariates in the data could take. This remained true despite the total number of observations being slightly more than 200 000 and an exact matching procedure ended up dropping more than 98% of all observations. Instead, a subclassification procedure was used that created 6 subclasses, in each of which the distribution of covariates were as similar as possible.

4.2 The Weibull model

The Weibull model is a parametric survival model that includes accelerated failure-time, meaning the survival times are examined with a log-linear model. This results in the treatment effect parameter being an expression of relative increase and decrease in survival time (Carroll, 2003).

This account of the Weibull model follows the description of the model by Carroll (2003). The following equation describes the probability density function of the Weibull model

$$f(t) = \alpha\lambda t^{\alpha-1}e^{-\lambda t^{\alpha}} \qquad (4.2.1)$$

Where *t* represents the time-to-event variable, $\alpha > 0$ is the shape parameter and $\lambda > 0$ is the event rate parameter. An alternative parameterization is given by setting

$$\alpha = \frac{1}{\sigma} \text{ and } \lambda_i = e^{\frac{-(\mu+\underline{\beta}'\underline{x_i})}{\sigma}} \qquad (4.2.2)$$

Where $\underline{x_i}$ is the influence of a given covariate on individual *i*, modelled with λ_i. With equation 4.2.1 we can get the hazard ratio for two treatments

$$\theta(t) = \frac{\alpha_A\lambda_A}{\alpha_B\lambda_B}t^{\alpha A-\alpha B} \qquad (4.2.3)$$

While we can use equation 4.2.2 to get an equation showing the log hazard ratio for individual *i* with covariates $\underline{x_i}$ relative to the log hazard ratio for individual *j* with covariates $\underline{x_j}$:

$$\frac{-\underline{\beta}'(\underline{x_i}-\underline{x_j})}{\sigma} \qquad (4.2.)$$

5 Analysis and results

To develop the analysis, the restricted data set was matched with the MatchIt procedure described previously – this is done for every dataset that is used for analysis in this thesis. Practically this is done with an equation where the treatment variable "Graded" is a function of all the other explanatory variables. While a couple of functional forms for the regression analysis was considered, the final form used for the survival analysis ended up including every relevant explanatory variable available from the dataset and with a three-way interaction term between gender, sector and the dichotomous dummy variable indicating whether the sick leave was graded or not (just named "Graded" in the analysis):

$$\begin{aligned} Surv(length) = {} & Gender * Sector * Graded + Number\ of\ sick\ leaves \\ & + Days\ since\ last\ sick\ leave + Occupation + Age\ group \end{aligned}$$

The interaction term was chosen to include three variables instead of the much more common two variable max because of the limited number of factors in each variable (gender and graded having the binary 0/1 values for female/male and ungraded/graded respectively, as well as Sector including only three factors: private, municipality and state). Additionally, while the inclusion of the graded interaction variable is obvious given that that's the treatment variable we're interested in exploring, we can expect gender to play a huge part in how long a sick leave is given the previous discussion of empirics of sickness absences and the inclusion of pregnancy-related absences in the data. We also expect the sector to play a large role in how sickness absences and leaves develop over time, because different sectors have different expectations and requirements from the employee point of view – informally, the view that a government job is a relatively "safe" job while a private sector job is much more competitive and prone to pruning/replacement is widespread in society.

The results of this regression using two datasets, one that removes any sick leave 14 days and shorter per Nossen & Brage (2013)'s suggestion (result (1), on the left side) and another dataset that removes any sick leave 56 days or shorter, equivalent to the data restriction that has been done in some of the literature described in the literature review (result (2), right side)

For all the results presented in this section, the reference levels are female, municipal sector, ungraded, occupation 0 (military/unspecified), age 20-24. The standard error for each estimation is included in parenthesis under the relevant estimated parameter value.

Table 4: regression results, actual sick leave days

	Dependent variable: Sick leave length	
	(1)	(2)
Male	0.041**	0.041
	(0.019)	(0.027)
Private sector	0.149***	0.049***
	(0.013)	(0.017)
State sector	-0.019	-0.004
	(0.014)	(0.021)
Graded	0.717***	0.115***
	(0.011)	(0.013)
Number of sick leaves	-0.017***	-0.009***
	(0.0005)	(0.001)
Days since last sick leave	-0.0002***	-0.00004***
	(0.00001)	(0.00001)
Occupation 1 (Legislators/leaders)	0.063**	0.036
	(0.029)	(0.032)
Occupation 2 (Professionals)	0.069**	0.057*
	(0.027)	(0.030)
Occupation 3 (Technicians/associate professionals)	0.093***	0.060*
	(0.027)	(0.031)
Occupation 4 (Clerks)	0.068**	0.073**
	(0.030)	(0.033)
Occupation 5 (Service/shop/market sales workers)	0.162***	0.106***
	(0.027)	(0.030)
Occupation 6 (Agricultural/fishery workers)	0.054	0.082
	(0.047)	(0.053)
Occupation 7 (Craft/related trades workers)	0.168***	0.114***
	(0.030)	(0.033)
Occupation 8 (Plant and machine operators/assemblers)	0.206***	0.097***
	(0.030)	(0.033)
Occupation 9 (Elementary occupations)	0.203***	0.128***
	(0.029)	(0.032)
Age group 2 (25-29)	0.080***	0.039**
	(0.016)	(0.017)
Age group 3 (30-34)	0.091***	0.076***
	(0.016)	(0.017)
Age group 4 (35-39)	0.127***	0.135***
	(0.016)	(0.017)
Age group 5 (40-44)	0.151***	0.197***
	(0.016)	(0.017)
Age group 6 (45-49)	0.183***	0.220***
	(0.016)	(0.017)
Age group 7 (50-54)	0.210***	0.236***
	(0.016)	(0.017)
Age group 8 (55-59)	0.248***	0.268***
	(0.016)	(0.017)
Age group 9 (60-64)	0.489***	0.402***
	(0.016)	(0.017)
Male * Private sector	-0.063***	-0.071**
	(0.023)	(0.031)
Male * State sector	-0.060**	-0.073*
	(0.028)	(0.040)
Male * Graded	-0.048*	-0.045
	(0.028)	(0.032)
Private Sector * Graded	-0.134***	-0.026
	(0.017)	(0.019)
State sector * Graded	-0.033*	-0.017
	(0.018)	(0.023)

Male * Private sector * Graded	0.075**	0.068*
	(0.034)	(0.037)
Male * State sector * Graded	0.066	0.105**
	(0.041)	(0.048)
Constant	3.847***	4.857***
	(0.031)	(0.035)
Observations	85,943	33,908
Log Likelihood	-452,146.200	-195,483.600
chi^2 (df = 30)	13,550.420***	1,883.059***
Note:		*p<0.1; **p<0.05; ***p<0.01

The first parameters of note are the Graded parameters. For the dataset where length > 2 weeks, the significant parameter of 0.717 indicates that using graded sick leaves contributes to a large increase in the length of the sick leave; in fact, it is the largest single parameter in absolute value in the results. Looking at how a gender and sector swap changes the results, they are all relatively small negative modifications to the large parameter, with private sector contributing the most to lowering the sick leave length – for males, a reduction of -0.063 (of the parameter for just Graded) and for females, a reduction of -0.134. Employment in the state sector also contributes to a smaller increase in length, while being male with a graded sick leave lowers the survival time of the sick leave compared to women, possibly because of the exclusion of pregnancy-related absences. The overall effect for the dataset where length > 14 days remains that using a graded sick leave is a large factor in increasing the length of the sick leave. A few of the interaction variables are of varying significance: Male * State sector * Graded is insignificant, Private sector * Graded and Male * Graded are only significant for $\alpha < 0.05$ and Male * Private Sector * Graded is only significant for $\alpha < 0.1$. α as it's used here is a measure of the chosen significance level for the results.

This result is a clear outlier when compared to previous studies. The likely cause for this is, assuming that the region of Troms is not unique in their work and sick leave dynamics, limitation of the data. Specifically, the lack of diagnosis data hurts the analysis in this regard, as discussed previously. Assuming that the results are accurate, the indications would be that for sick leaves longer than 14 days the lock-in effect and the effect of returning to work too early and/or with too low grading dominates over the positive benefits described in the introduction.

The second set of regression results, (2) in table 4, includes data where sick leave > 56 days, or 8 weeks. By imposing such a harsh restriction on the dataset, we hope to filter out a lot of contagious diseases, pregnancy-related absences and shorter-term illnesses that tends to fully incapacitate a person that typically lasts longer than 14 days. It also limits the analysis to the individuals with very long-term absences rather than the long-term and medium-term absences in (1), meaning the results could have interesting implications for that very specific group of people.

The new, still significant, Graded parameter is 0.115 which is a large decrease from the previous dataset's regression results of 0.717 – a decrease of almost 84%. The interaction term parameters are mostly insignificant and either way they are all very small, notably Male * Private sector * Graded has a parameter value of 0.068 that is significant for $\alpha < 0.1$ and Male * State sector * Graded has a parameter value of 0.105 significant for $\alpha < 0.05$. This indicates that being a male in both state and private sector with a graded sick leave contributes to a slight increase in the length of the sick leave, more so for state sector than private.

These results are more closely aligned with previous knowledge in the field, although it still suggests that using graded sick leaves is an overall contributor to longer sick leaves rather than shorter, even if the effects are significantly reduced by excluding any sick leave 8 weeks or shorter. The negative effects that dominated for sick leaves longer than 14 days still dominate the positive benefits.

In the introduction I briefly discussed the idea of effective sick leave length – that is, measuring the total amount of productivity lost rather than measuring the amount of calendar days your sick leave lasts for. From a societal point of view, if your goal is to extract productivity from people with sickness absences, it is preferable to have a person on a 40% graded sick leave for 10 days than a 100% graded sick leave for 5 days – obviously, this is only preferable in a very limited model that assumes such a pattern won't affect future sick leaves, and completely ignores any other negative effect that might arise from extracting productivity at the cost of recovery time. To incorporate this idea into the regression model, the dependent variable is changed from just length to length multiplied by average grading – thereby changing it to the effective sick leave. The explanatory variables remained unchanged from the previous analysis, and the new regression has the functional form

$$\begin{aligned} &Surv(length * average\ grading) \\ &\qquad = Gender * Sector * Graded + Number\ of\ sick\ leaves \\ &\qquad + Days\ since\ last\ sick\ leave + Occupation + Age\ group \end{aligned}$$

Mirroring the first set of results, the analysis is done with two sets of data, one that excludes any sick leave with a length of 14 days and shorter and one that excludes any sick leave with a length of 56 days or shorter. The regression results with these data sets are presented in column (1) and (2), respectively:

Table 5: regression results, effective sick leave days

	Dependent variable:	
	Sick leave length * Average grading	
	(1)	(2)
Male	0.048**	0.055*
	(0.020)	(0.030)
Private sector	0.142***	0.039**
	(0.014)	(0.019)
State sector	-0.007	0.011
	(0.015)	(0.024)
Graded	0.312***	-0.247***
	(0.011)	(0.015)
Number of sick leaves	-0.016***	-0.009***
	(0.0005)	(0.001)
Days since last sick leave	-0.0001***	-0.00004***
	(0.00001)	(0.00001)
Occupation 1 (Legislators/leaders)	0.054*	0.030
	(0.031)	(0.037)
Occupation 2 (Professionals)	0.055*	0.041
	(0.029)	(0.035)
Occupation 3 (Technicians/associate professionals)	0.088***	0.054
	(0.029)	(0.035)
Occupation 4 (Clerks)	0.070**	0.075**
	(0.031)	(0.037)
Occupation 5 (Service/shop/market sales workers)	0.183***	0.134***
	(0.029)	(0.035)
Occupation 6 (Agricultural/fishery workers)	0.087*	0.131**
	(0.050)	(0.061)
Occupation 7 (Craft/related trades workers)	0.184***	0.136***
	(0.032)	(0.037)
Occupation 8 (Plant and machine operators/assemblers)	0.231***	0.131***
	(0.032)	(0.038)
Occupation 9 (Elementary occupations)	0.243***	0.189***
	(0.031)	(0.037)
Age group 2 (25-29)	0.082***	0.038*
	(0.017)	(0.020)
Age group 3 (30-34)	0.081***	0.057***
	(0.017)	(0.019)
Age group 4 (35-39)	0.110***	0.104***
	(0.017)	(0.019)
Age group 5 (40-44)	0.123***	0.153***
	(0.017)	(0.019)
Age group 6 (45-49)	0.160***	0.179***
	(0.017)	(0.019)
Age group 7 (50-54)	0.180***	0.188***
	(0.017)	(0.019)
Age group 8 (55-59)	0.216***	0.212***
	(0.017)	(0.019)
Age group 9 (60-64)	0.467***	0.372***
	(0.017)	(0.019)
Male * Private sector	-0.067***	-0.079**

	(0.025)	(0.035)
Male * State sector	-0.066**	-0.087*
	(0.030)	(0.046)
Male * Graded	-0.029	-0.035
	(0.030)	(0.036)
Private Sector * Graded	-0.104***	0.004
	(0.018)	(0.022)
State sector * Graded	-0.037*	-0.007
	(0.020)	(0.026)
Male * Private sector * Graded	0.110***	0.102**
	(0.036)	(0.042)
Male * State sector * Graded	0.077*	0.122**
	(0.044)	(0.054)
Constant	3.820***	4.830***
	(0.033)	(0.040)
Observations	85,943	33,908
Log Likelihood	-435,792.700	-189,021.800
chi^2 (df = 30)	4,835.311***	2,780.829***
Note:		*p<0.1; **p<0.05; ***p<0.01

Focusing on the first set of regression results first (sick leave > 14 days), the Graded parameter is 0.312 and significant, indicating that using graded sick leaves for this dataset increases the length of the sick leave to a not insignificant degree. Graded sick leaves in the state sector reduces this by a little bit, -0.037 significant for $\alpha < 0.1$ while graded sick leaves in the private sector actually reduces it by quite a bit with a large significance, -0.104. The interpretation is that graded sick leaves have a smaller negative effect on leave length (negative as in it increases leave length) when used in the private sector than in other sectors for females. Being a male in private or state sector with a graded sick leave contributes to a larger increase in effective leave length compared to the equivalent female. Overall the effects of graded sick leaves on effective leave length are that they increase the effective leave length, meaning they are still a net loss productivity-wise to society. These results more or less mirror what we saw in the actual sick leave length with the same dataset, just with a smaller effect. Measuring effective leave length rather than actual leave length therefore gives a more accurate view of the effects on overall productivity in society, which are smaller than what just looking at actual leave length would otherwise lead you to believe.

The really interesting results come from the effective leave length analysis with the dataset that excludes any leave 8 weeks or shorter. We see a flip to a negative parameter for the Graded variable, indicating that in this regression, applying graded sick leaves contributes to a shortening of effective leave length – a positive result. The rest of the interaction variables are mostly insignificant when it comes to graded sick leave. Male * Private sector * Graded and Male * State sector * Graded are significant for $\alpha < 0.05$ and positive of roughly half the value of the graded parameter's absolute value. The overall effect is still a shortening of effective leave length when using graded, but for males in private and state sector the positive benefit is roughly half than that of the reference female, municipality sector.

6 Conclusions

To properly draw solid conclusions from the results presented in this thesis, the analysis should be redone with diagnosis data included, as I believe the lack of such a critical explanatory variable and data filtering tool really hurts the overall analysis. Nevertheless, the results do lend themselves to some concluding remarks.

In general, the analysis shows that for Troms county the effect of using graded sick leaves are an overall increase in the length of the sick leave. These effects become smaller as the leave length that receives grading increases, indicating that the negative lock-in effects and risk of returning to work too early either decrease as sick leave length increases or that the positive effects of activity related to workplace participation during sick leaves described in the introduction increase and begin to dominate as sick leave length increases. Intuitively, one might expect that the lock-in effect actually increases as leave length increases assuming that people become comfortable with their current work arrangement and see a potential upcoming increase in workload when a sick leave ends as a negative. The other risk mentioned, that of overworking/returning to work too early, is unclear as to how it might develop as leave length increases.

The positive effects intuitively make sense to dominate as leave length increases. Since they have to do with activity and connection to the workplace, you would expect them to have larger effects if they are compared to the alternative of long-term inactivity and disconnection from the workplace – put differently, the cost of not receiving a graded sick leave when an individual is placed on a long-term sick leave stacks up quickly. This thesis does not do any formal analysis on isolating the different effects of graded sick leave, but this could be an interesting avenue of future research in the field.

What we do see however is that the effective sick leave, a better measure of actual productivity lost than just sick leave days, decreases when applied to only very long-term sick leaves of longer than 8 weeks. While no other of the reviewed literature attempted to make a distinction between actual sick leave length and effective sick leave length, these results are in line with the general results in the field of graded sick leaves leading to a lowered sick leave length. With the aforementioned improved data the positive results may extend to the actual sick leave measures as well.

In conclusion, one should be careful to extend the results found here regarding actual sick leave length to conclude that previous research is wrong. The policy implications of the results are that there could be benefits to intensifying the use of graded sick leaves for individuals with very long sick leaves, however more research should be done on consequences of grading sick leaves other than the effects on leave length before any solid recommendation can be made.

7 References

12452: Doctor certified sickness absence for employees (per cent), by occupation, quarter, sex and contents. (n.d.). Retrieved May 27, 2019, from https://www.ssb.no/en/statbank/table/12452/tableViewLayout1/

Andrén, D. (2014). Does part-time sick leave help individuals with mental disorders recover lost work capacity? *Journal of Occupational Rehabilitation*, *24*(2), 344–360. https://doi.org/10.1007/s10926-013-9467-4

Atle Lie, S. (2014). *Liten effekt av gradert sykmelding for raskere retur til arbeid: En kausal analyse av registerdata for varighet av sykefravær, ulike trygdeytelser og retur til jobb.* Bergen.

Berge, C. (2012). Internasjonal sammenligning av sykefravær. *Økonomiske Analyser*, (2), 6.

Brage, S., Kann, I. C., Kolstad, A., Nossen, J. P., & Thune, O. (2011). Gradert sykmelding - omfang, utvikling og bruk. *Arbeid Og Velferd*, (3), 32–44.

Bruusgaard, D., & Claussen, B. (2010). Ulike typer sykefravær. *Tidsskrift for Den Norske Laegeforening*, *130*(18), 1834–1836. https://doi.org/10.4045/tidsskr.10.0541

Carroll, K. J. (2003). On the use and utility of the Weibull model in the analysis of survival data. *Controlled Clinical Trials*, *24*(6), 682–701. https://doi.org/10.1016/S0197-

2456(03)00072-2

Croson, R., & Gächter, S. (2010). The science of experimental economics. *Journal of Economic Behavior and Organization*, *73*(1), 122–131. https://doi.org/10.1016/j.jebo.2009.09.008

Egenmelding. (2019). Retrieved May 27, 2019, from https://www.nav.no/no/Person/Arbeid/Sykmeldt%2C+arbeidsavklaringspenger+og+yrkesskade/Relatert+informasjon/egenmelding

Gradert sykemelding. (n.d.). Retrieved May 27, 2019, from https://arbinn.nho.no/arbeidsrett/sykefravar_og_permisjoner/sykefravarsoppfolging/artikler/gradert-sykemelding/

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Ssrn*. https://doi.org/10.1093/pan/mpl013

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). **MatchIt** : Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, *42*(8). https://doi.org/10.18637/jss.v042.i08

Høgelund, J., Holm, A., & McIntosh, J. (2010). Does graded return-to-work improve sick-listed workers' chance of returning to regular working hours? *Journal of Health Economics*, *29*(1), 158–169. https://doi.org/10.1016/j.jhealeco.2009.11.009

Markussen, S., Mykletun, A., & Røed, K. (2012). The case for presenteeism - Evidence from Norway's sickness insurance program. *Journal of Public Economics*, *96*(11–12), 959–972. https://doi.org/10.1016/j.jpubeco.2012.08.008

Meneses-Echavez, J. F., Baiju, N., & Berg, R. C. (2018). *Effects of partial sick leave versus full-time sick leave on sickness absence and work participation: a systematic mapping review*. Oslo.

Moore, D. F. (2016). *Applied Survival Analysis Using R*. https://doi.org/10.1007/978-3-319-31245-3

Nossen, J. P., & Brage, S. (2013). Effekt av gradert sykmelding - betydningen av metodevalg.

Arbeid Og Velferd, (3), 68–79.

Standard for yrkesklassifisering (STYRK-08). (2011).

Viikari-Juntura, E., Kausto, J., Shiri, R., Kaila-Kangas, L., Takala, E.-P., Karppinen, J., … Martimo, K.-P. (2012). Return to work after early part-time sick leave due to musculoskeletal disorders: a randomized controlled trial. *Scandinavian Journal of Work, Environment and Health*, *38*(2), 134–143. Retrieved from http://media.proquest.com/media/pq/classic/doc/2617423041/fmt/pi/rep/NONE?hl=&cit%3Aauth=Viikari-Juntura%2C+Eira%2C+MD%2C+PhD%3BKausto%2C+Johanna%2C+DDS%2C+MSocSc%3BShiri%2C+Rahman%2C+MD%2C+PhD%3BKaila-Kangas%2C+Leena%2C+PhD%3BTakala%2C+Esa-Pekka%2C+MD%2C

Waddell, G., & Burton, A. K. (2006). *Is work good for your health and well-being?* The Stationary Office.

8 Appendix

8.1 Occupation codes

Information on occupation codes retrieved from (*Standard for yrkesklassifisering (STYRK-08)*, 2011). Descriptions are included where the title is not self-explanatory.

Occupation 0: Armed forces and unspecified
Includes any occupation that fits the description of "military personnel", as well as any occupation that can't be identified due to inadequate or missing information.

Occupation 1: Politicians and leaders
Includes all types of political work, as well as any managerial position in private, state and municipality sector.

Occupation 2: Professionals
Includes any occupation that would typically require a higher education of a minimum length of 4 years

Occupation 3: Technicians and associate professionals
Includes any occupation that would typically require a higher education with a length of 1-3 years.

Occupation 4: Clerical support workers

Occupation 5: Service and sales workers

Occupation 6: Skilled agricultural and fishery workers

Occupation 7: Craft and related trade workers

Occupation 8: Plant and machine operators and assemblers

Occupation 9: Elementary occupations
Includes any occupation that is generally considered unskilled, including cleaning, unskilled labourers, refuse workers and others.

8.2 Age groups

Table 6: age groups

Age group	Age range in years
1	20-24
2	25-29
3	30-34
4	35-39
5	40-44
6	45-49
7	50-54
8	55-59
9	60-64