Social Interactions and Sickness Absence:

Family, Colleague and Neighborhood Effects

Universitetet
i Stavanger

Gunn Halvorsen

Universitetet
i Stavanger

DET SAMFUNNSVITENSKAPELIGE FAKULTET,
HANDELSHØGSKOLEN VED UIS

MASTEROPPGAVE

STUDIEPROGRAM:	OPPGAVEN ER SKREVET INNEN FØLGENDE SPESIALISERINGSRETNING:
Master i økonomi og administrasjon	Økonomisk analyse
	ER OPPGAVEN KONFIDENSIELL? Nei (**NB**! Bruk rødt skjema ved konfidensiell oppgave)

TITTEL:

Sosiale interaksjoner og sykefravær: Familie, kollega og nabolagseffekter

ENGELSK TITTEL:

Social interactions and sickness absence: Family, colleague and neighborhood effects

FORFATTER(E)		VEILEDER:
Studentnummer:	Navn:	Professor Kenneth H. Wathne
204627	Gunn Halvorsen	Professor Mari Rege

OPPGAVEN ER MOTTATT I TO – 2 – INNBUNDNE EKSEMPLARER

Stavanger, ……/…… 2013 Underskrift administrasjon:……………………………

Acknowledgements

This thesis represents the final part of a two year master's degree program in Business Administration at the University of Stavanger (UiS).

First and foremost I would like to thank my thesis advisors professor Kenneth H. Wathne and professor Mari Rege for extraordinary engagement, for always being available, and for all help and support. In addition I would like to give a special thanks to Auke Hunneman for all his time and assistance, making it possible for me to include spatial analysis.

Finally, I would like to thank Venke F. Haaland for her good assistance whenever needed.

Gunn Halvorsen

Stavanger, June 2013

Abstract

Social interactions play an important role in economic decision making. In this thesis I examine the effects of social influence on sickness absence. While I study social interaction effects among i) family members, ii) colleagues, and iii) neighbors, the main focus are on neighbors.

I use Norwegian data provided by Statistics Norway to analyse if the percentage of one's family members, colleagues, and neighbors on sick leave affects an individual's likelihood of being on sick leave. The results indicate presence of social interaction effects in all three groups. The estimated family effect suggests an increased sick leave probability of 0.07 percent if the percentage of one's family members on sick leave increases with one percent. Individual's faced with a one percent increase in the percentage of his or her colleagues on sick leave, show a 0.28 percent increased probability of receiving sick pay. Finally, the estimated neighborhood effect show that a one percent increases in the percentage of neighbors on sick leave show an increased sick leave probability of 0.39 percent.

While I do find support for social interaction effects across all models, I acknowledge the methodological challenges related to estimating group effects using the above approach. To address one of these challenges, I estimate a neighborhood model that accounts for the so-called reflection problem. The result from this model is consistent with the simpler model, providing further evidence of social influence on individual sickness absence by neighbors. Lastly, I estimate a spatial model to assess the impact by nearby geographical neighborhoods on sickness absence in a focal neighborhood. The result indicates that the percentage of individuals on sick leave in a focal neighborhood is influenced by these neighboring neighborhoods.

Table of contents

Tables, Figures and Appendix

1. Introduction

In spite of generally good health and long life expectancy among Norwegians, the costs related to sick leave and early retirement in Norway are double that of the OECD country average based on GDP (OECD, 2012).

Identifying the determinants of sickness absence has been highly researched. Economic studies mostly focus on economic incentives (e.g., Johansson & Palme, 1996), workplace characteristics (e.g., Arai & Thoursie, 2004), and macroeconomic fluctuations (e.g., Arai & Thoursie, 2005). While these determinants have been found important to the fluctuations in sickness absence, they cannot fully explain differences in patterns of sickness absence across and within regions (Bokenblom & Ekblad, 2007). It has also proved difficult to explain the observed patterns by benefit rules, socioeconomic factors and general (measurable) health conditions (Lindbeck, Palme, & Persson, 2007). Thus, it has been argued that observed patterns in sickness absence could be related to social interaction, such as group effects on individual behavior (Lindbeck et al., 2007).

Empirical estimation of social interaction effects often rely on estimating how the behavior of the individual are affected by the average behavior of the individual's reference group (Bokenblom & Ekblad, 2007). The results from existing studies do indicate that the average behavior of the group is often important in explaining individual sickness absence (e.g., Bokenblom & Ekblad, 2007; Dale-Olsen, Nilsen, & Schøne, 2010; Lindbeck et al., 2007).

The purpose of this thesis is also to investigate if social interaction effects are important in understanding individual sickness absence. I study social interaction effects among i) neighbors, ii) family members, and iii) colleagues. First I replicate parts of the study by Lindbeck et al. (2007) which examines social interaction effects within neighborhoods. Second, I examine family effects, and third I use Dale-Olsen et al.'s (2010) study as a benchmark to model the effects of colleagues on sickness absence. Finally, I apply spatial econometrics to more closely examine neighborhood effects. Although I examine social interaction effects in three different groups, one of which has not previously been examined (i.e. family), it will become clear throughout the thesis that the main focus is on social interaction effects among neighbors.

The data set used in this thesis is obtained from a database called *FD-trygd* provided by Statistics Norway. This database contains detailed records for every Norwegian from 1992 to 2003. My empirical analysis mostly relies on data from 2003. This sample contains 1,527,533 workers, living in 13,123 different neighborhoods, working in 14 different industries. The data set only reports people that have been on sick pay for more than 16 days. Employers cover compensation for shorter durations, and thus are not systematically reported at the individual level. The estimates reported in this thesis should thus be interpreted as the probability of receiving sick pay above 16 days.

My results show indication of social interaction effects among all the three groups that I study. Results obtained from estimating family effects suggest an increased sick leave probability of 0.07 percent if the percentage of one's family members on sick leave increases with one percent. Individuals faced with a one percent increase in the percentage of his or her colleagues on sick leave, show a 0.28 percent increased probability of receiving sick pay. Finally, a one percent increases in the percentage of neighbors on sick leave show an increased sick leave probability of 0.39 percent. Moreover, using spatial analysis at the neighborhood level, I find that the percentage of individuals on sick leave in nearby geographical neighborhoods affects sickness absence in a focal neighborhood.

There are many well-known methodological challenges related to estimating group effects. I control for one of these, the reflection problem, by adopting an approach used by Lindbeck et al. (2007), where they examine how immigrants are affected by the sickness absence of natives living in the neighborhood where they settle after arriving in Sweden. The estimated effect for this model suggest that if the percentage of Norwegians on sick leave in a neighborhood increase with one percent, an immigrant's probability of being on sick leave increases with 0.24 percent.

The reminder of this thesis is organized as follows: In section two existing research into the effects of social interaction on sickness absence is reviewed. In section three I present some of the concerns that have been raised concerning measuring social interaction effects. In section four I provide a review of the sickness benefit system in Norway, and characteristics of the Norwegian labor market. This is followed by a presentation of the data set in section five. I then move on to present the empirical strategy and results in section six. Finally, I draw conclusions based on the findings and discuss their implications in section seven. In the final section I also discuss limitations of the current study and directions for further research.

2. Existing Literature

According to Dale-Olsen et al. (2010) the importance of social influence and norms for absenteeism has been studied in the social psychology and management literatures at least since Chadwick-Jones, Nicholson, and Brown (1982) documented the large variance in absence across industries, organizations, and intra-organizational units.

In the economic literature there is a large and growing body of research focusing on the importance of social interaction on a wide variety of output variables like; educational choices (Lalive & Cattaneo, 2009), labor supply (Grodner & Kniesner, 2006), retirement decisions (Duflo & Saez, 2003), and disability behavior (Rege, Telle, & Votruba, 2012).

This thesis relates directly to the economic literature by focusing on the role of social interaction effects on sickness absence. The literature on social interaction effects and sickness absence is small but growing. Dale-Olsen et al. (2010) use Norwegian data to examine how colleagues affect each other's work absence. They use two measures of sickness absence i) the number of sickness absence spells, and ii) the duration of sickness absence. Their preferred estimates suggest that social interaction effects at the workplace do exist, and that the effects are noticeable in size. However they also recognize that, even after controlling for endogeneity as well as including a large battery of control variables, they cannot draw unambiguous conclusions about the exact nature of the relationships they present. E.g., the results could still be partly driven by workload caused by absence from other colleagues.

Rieck, Marshall, and Vaage (2012) investigate whether teachers' sickness absence is affected by the average absence level of fellow teachers, using Norwegian data. They adopt different approaches where methodological problems such as the reflection problem and intra-group correlation are mitigated. Their results show that the significance of the social interaction effects critically depends on their ability to control for unobserved school characteristics.

Bokenblom and Ekblad (2007) use Swedish data to analyse social interaction effects among colleagues. They estimate to what extent an individual's share of sickness absence is influenced by the share of sickness absence of the worker's colleagues. Their results show positive and significant peer group effects. Furthermore, their results are robust over different model specifications and qualitatively equivalent when different estimations strategies are applied. They find that, on average, it takes two to three years for a new employee to adopt the pattern of the work group. Lastly, they find evidence of peer effects to be an intra-gender and intra-age group phenomenon which strengthens their belief that the peer effect reflects some form of social phenomenon.

Lindbeck et al. (2007) use Swedish data to analyse if the average level of sickness absence in a neighborhood affect individual sickness absence through social interactions. They adopt several different approaches to deal with well-known methodological problems. Their results are robust in the sense that regardless of approach and identifying assumption, the estimated group effects are statistically significant.

A closely related study to the context that I examine in this thesis, is the recent study conducted by Markussen and Røed (2012) on administrative panel data from Norway. They use a fixed-effects methodology to examine how *social insurance dependency* spreads within neighborhoods, families, ethnic minorities, and former schoolmates. Their estimated network effects are both quantitatively and statistically significant, and rise rapidly with "relational closeness" in a way that establishes endogenous social interaction as a central causal mechanism. They find no evidence that social interactions cross ethnic borders.

3. Measuring Social Influence

As evident from the existing literature, empirical estimation of social interactions often rely on estimating how the behavior of the individual is affected by the average behavior of the individual's reference group. There are several methodological challenges related to the estimation of group effects using this approach (Lindbeck et al., 2007). In the following I review some of these methodological challenges, starting with homophily (or "correlated effects").

3.1. Homophily

"Homophily is the principle that a contact between similar people occurs at a higher rate than among dissimilar people" (McPherson, Smith-Lovin, & Cook, 2001, p. 416). Initial studies on informal network ties showed substantial homophily in terms of demographic characteristics such as age, sex, ethnicity, education, and psychological characteristics often linked to demographic characteristics (McPherson et al., 2001). Ever since, researchers have studied homophily over a wide range of different relations from marriage and friendships, to appearing with someone in a public space (McPherson et al., 2001). Despite some subtle differences, in general the patterns of homophily are remarkably robust (McPherson et al., 2001).

Broadly speaking we can identify two theoretically distinct mechanism by which homophily arises – *choice homophily* and *induced homophily* (Kossinets & Watts, 2009). The former is where homophilous ties can be attributed to individual, psychological preferences, and the latter emerges as a consequence of the homogeneity of structural opportunities for interaction, like at a workplace (Kossinets & Watts, 2009).

"While geography is the physical substrate on which homophily is built, family connections are the biosocial web that connect us to those who are simultaneously similar and different" (McPherson et al., 2001, p. 431). While family ties have a somewhat different structure than more voluntary and less intense social ties like co-employment or friendship, there are still fundamental similarities (McPherson et al., 2001).

Ties formed among co-workers are generally more heterogeneous in race and religion than ties formed elsewhere, and more homogeneous on sex and education. "Employees are especially likely to have ties to others who occupy their same job, and occupational sex segregation induces strong baseline homophily" (McPherson et al., 2001, p. 434). Educational and occupational homophily has been documented by researchers in a large number of societies, although there are some indications that its level varies somewhat from country to country (McPherson et al., 2001).

Because the homophily principle structures network ties of every type, including family, work, and neighborhood, the result is that people's personal networks are homogenous with respect to many socio-demographic, behavioral and interpersonal characteristics (McPherson et al., 2001).

Based on the well-documented homophily principle, it's clear that when studying social interactions, it's important to be cautious when interpreting the results. E.g., if it is found that the percentage of people in a neighborhood on sick pay affect individual's sickness absence, this could be due to homophily. It is therefore important to try to control for such effects. Aral, Muchnik, and Sundararajan (2009) present a generalized statistical framework for distinguishing peer-to-peer influence from homophily in dynamic networks of any size. When accounting for homophily and individual characteristics their estimates of influence are substantially reduced (Aral et al., 2009).

3.2. The Reflection Problem

Two hypotheses often advanced to explain the common observation that individuals belonging to the same group tend to behave similarly, are endogenous and correlated effects (Manski, 2003). Endogenous effects refer to the propensity for an individual's behavior to vary with the behavior of the group (e.g., pressure to conform to group norms) .Whereas correlated effects is the tendency for people with similar unobserved characteristics to be in the same group (homophily), or being exposed to similar local differences (common cause).

Trying to estimate social interactions based on how the behavior of the individual is affected by the average behavior of the individual's reference group, give raise to a simultaneity problem. This simultaneity problem is what Manski (1993) defines as the "reflection" problem – group behavior affects individual *i*'s behavior, which in turn, affects the groups behavior. That is, we assume that the individual is affected by the average sickness absence of the group, but per definition, the average sickness absence under this assumption will then also be affected by the sickness absence of that same individual.[1] Thus, the reflection problem hinders the identification of the endogenous from correlated effects (Bramoullé, Djebbari, & Fortin, 2009).

Policy implications are the main reason why one is generally concerned about separating these two effects. Only endogenous effects have the potential to create so called "social multipliers" (Manski, 1993). Social multipliers would, in the context of this thesis, imply that if one where able to reduce sickness absence within a specific group, one would not only benefit from the specific cases, but also get a positive indirect effect on the rest of the group.

According to Manski (2003) even innumerable empirical observations of the behavior of individuals in groups can't *per se* distinguish between these effects. "To draw conclusions requires that empirical evidence be combined with sufficiently strong maintained assumptions about the nature of individual behavior and social interactions" (Manski, 2003, p. 12). Bramoullé et al. (2009) provide new results regarding the identification of peer effect by introducing an extended version of the linear-in-means model where interactions are structured through a social network. They assume that correlated unobservables are either

[1] This is also the case even when we exclude the individual from the average sickness absence in the group before estimating the effect.

absent, or treated as network fixed effects, and they provide an easy-to-check necessary and sufficient conditions for identification (Bramoullé et al., 2009).

3.3. Unobserved heterogeneity

Unlike the other methodological challenges, unobserved heterogeneity is a more general concern. It will however become clear that this is a special concern in my thesis, and thus addressed here.

Unobserved heterogeneity is a form of omitted variable bias, and refers to omitted variables that are fixed for an individual (or at least over a long time period) (Murray, 2006). With cross-sectional data, there is no particular reason to differentiate between omitted variables that are fixed, and omitted variables that change over time (Murray, 2006). Bias from unobserved heterogeneity is particularly important in non-linear regression models, because unlike linear regression models, estimated coefficients will be biased *even* if the unobserved heterogeneity is not correlated with the observed independent variables (Holm, Jæger, & Pedersen, 2008). When one has panel data, there are number of ways to address this potential problem (e.g., random effects). However, when one uses cross-sectional data as in this thesis, it's difficult to deal effectively with potential bias from unobserved heterogeneity because there is little information in the data that allows me to identify and correct the potential bias (Holm et al., 2008).

4. Institutional Details

4.1. The Norwegian Public Sickness Benefit System

Sickness benefits provide compensation for loss of income for workers who are unable to work due to illness or injury. In Norway you are entitled to sick pay from the first day of sickness if you are employed, and have been so for the last four weeks. The entitlement is limited to a maximum of one year. Workers who are not able to return to work after this period are offered rehabilitation and benefits to qualify for other types of work.

Spells of sickness up till three days is based on self-certification[2], while longer spells require certification by a physician.[3] Employers compensate workers the first 16 days of sickness

[2] Self-certification may be used for 24 calendar days during a 12-month period.

absence. After this period, sick pay is publicly disbursed and administrated by the Norwegian labor and welfare administration (NAV).

The part of the worker's income that entitles sick pay[4], have to be at least 50 percent of the basic amount to have right to sick pay from NAV.[5] For most workers the benefit level is set to 100 percent of previous earnings. Workers with labor income that exceeds six times the basic amount are not compensated for income above this threshold, though the majority of employers offer top up for these high income workers. For a detailed review of the Norwegian public benefit system I refer to chapter eight of Folketrygdloven (1997).

4.2. Characteristics of the Norwegian Labor Market

The Norwegian labor market is characterized by a high degree of organization, high contractual coverage, and a relatively highly coordinated and centralized wage determination (NOU 2011:1). In the following I review some other important characteristics of the Norwegian labor market.

4.2.1. Participation in the Labor Market

Participation in the labor market in Norway is high and among the highest in the OECD-area (NOU 2009: 10). The participation is highest among 30- and 40-year olds, and after the age 62 there is a great departure from the labor market (NOU 2010: 1).

Labor force participation increased dramatically since the 1970s, largely due to women's entry into the labor market (NOU 2009: 10). The gender differences in labor force participation have since been significantly reduced. In 2008 the difference in employment rates between women and men was about 6 percentage points, compared to over 30 percentage points in the early 1970s (NOU 2009: 10). The largest gender gap in terms of labor force participation is found in the age group 60-64 (NOU 2009: 10). Here, the participation rate is about 12 percentage points higher among men than among women (NOU 2009: 10). The fact that the differences increase with age reflect both differences in labor

---

[3] If the workplace is part of the "Intergraded working life (IW)"-treaty, workers can use self-certification up to eight calendar days. Employers are entitled to allow longer absence period than 3 days without being certified by a physician.

[4] The income base for sick pay is generally the work income he or she received immediately before the disability occurred.

[5] The basic amount is used as the basis for calculating Norwegian social security and pension benefits. By May 2012, the basic amount was equal to 82,122 NOK.

market attachment for different cohorts, and that more women than men leave the workforce after aged 50 (NOU 2009: 10).

Labor force participation increases with education level. People with university or college degrees has about 30 percentage points higher employment rates than those with lower secondary school as their highest education (NOU 2009: 10). The difference increases with age, reflecting that those with low educational level leave the labor market earlier. Educational importance is somewhat greater for women than for men. For women the difference is 33 percentage points, while the difference is 26 percentage points among men (NOU 2009: 10). The increased level of education in recent decades has significantly contributed to reduced wage inequality between the genders (NOU 2009: 10).

4.2.2. Part Time Workers

A special feature of the Norwegian labor market is the high levels of part-time work. Data from the European Working Conditions Survey (EWCS) shows that while 27 percent of Norwegian employees report that they work part-time, the general corresponding figures for EU were 17 percent (NOU 2010: 1).[6] Like all the other countries in EU/EEA, women make up the largest proportion of part-time workers (NOU 2010: 1). 41 percent of female employees in Norway work part-time, compared to 13 percent of male employees (Statistisk Sentralbyrå, 2013a).

4.2.3. Labor Immigration

Labor immigration to Norway varies from year to year. From 2004 until 2008 there were particularly high growths in labor immigration to Norway (NOU 2010: 1). Much of this increase was due to the expansion of EU that led to Norway being a part of a labor market with 500 million people (NOU 2010: 1). In parallel with the economic downturn in 2008, the immigration slowed down (NOU 2010: 1).

Norway has been a part of a common Nordic labor market since 1954 (NOU 2010: 1). This means that immigrants from other Nordic countries do not need a work permit to work in Norway, and thus not included in most statistics regarding labor immigration and work permits (NOU 2010: 1).

---

[6] Norway participated in this study in 2000 and 2005 as part of the EEA agreement.

4.2.4. Sickness Absence

As mentioned in the introduction, the direct costs related to sickness absence are very high in Norway. This is both a direct result of a generous sickness benefit system as well as high absence rates. From 2011 to 2012,[7] the sickness absence rate increased not only in all counties in Norway, but also in all sectors (Statistisk Sentralbyrå, 2013b).[8] The strongest increase was found in local government (Statistisk Sentralbyrå, 2013b). In addition, all industries except one faced increased sickness absences rates (Statistisk Sentralbyrå, 2013b).

The sickness absence rate increased both for men and women in this same period. The rate increased from 8.3% to 8.6% for women, whereas there was an increase from 5.0% to 5.2% for men (Statistisk Sentralbyrå, 2013b).[9] The age groups 45-49 and 50-54 had the highest increase in sickness absence reported by a physician (Statistisk Sentralbyrå, 2013b). The only group facing decreased sickness absence were men in the age 60-64 (Statistisk Sentralbyrå, 2013b).

5. Data Set Description

The data set used in this thesis is obtained from a database called *FD-trygd* provided by Statistics Norway. This database contains detailed records for every Norwegian from 1992 to 2003, including individual demographic information, socioeconomic data, current employment status, and geographic identifiers.

The data used in this thesis is cross-sectional data covering the time period 2000-2003. Except for section 6.5., all analysis relies on data from the last year of which I have observations, namely 2003.

The data set only reports individual's that have received sick pay longer than 16 days during the year. Employers cover compensation for shorter durations, and thus are not systematically reported at the individual level. The estimated probability of receiving sick pay reported in this thesis is therefore the probability of receiving sick pay above 16 days.

Even though the data set covers the entire population in Norway, I confine the study to workers in the age group 18-64. I exclude any persons identified in *FD-trygd* as self-

[7] Changes reported are one year from the 4th quarter of 2011 to the 4th quarter of 2012.
[8] Sickness absence rate measures days lost due to own illness as a percentage of scheduled man days.
[9] The sickness absence rates reported here are not seasonal or flu adjusted.

employed, or receivers of disability benefits. I use family, neighborhood and workplace identifications to match each individual in the sample to their group of belonging. I exclude individuals with less than five colleagues and individuals with more than one workplace. Applying these restrictions yields a sample for the 2003 data of 1,527,533 workers, living in 13,123 different neighborhoods, working in 14 different industries. Appendix 1 provides a list of the socioeconomic variables that are included in the models in this thesis, and Table 1 provides summary statistics of these.

6. Empirical Strategy and Results

In this section I start by examining social interaction effects among neighbors. Most models were first estimated as linear probability model's (LPM's).[10] As I will discuss there are several potentials challenges with this approach. All LPM's were thus subsequently estimated using logit regressions. I then compare all models represented thus far over time, to see if the results are consistent over time. Next I examine whether there is indications of social interaction effects among family members, before examining social interaction effects among colleagues. The section ends with a spatial analysis, to more closely examine neighborhood effects.

6.1. A First Look at the Data

To study neighborhood effects I use basic statistical units (*grunnkrets*) as the geographic unit of analysis. In Norway there are about 14,000 basic statistical units, which provide stable, coherent and generally homogeneous geographical units (Statistisk Sentralbyrå, 1999). In the following these will be referred to as "neighborhoods".

To assess local variation in sickness absence, I first examined the percentage of individuals that received sick pay across neighborhoods in 2003. Let S_{in} be the binary variable measuring if individual *i* living in neighborhood *n* received sick pay during the year (1 = received sick pay), $\bar{S}_n$ the average percentage of people that received sick pay in that neighborhood, and $\bar{\bar{S}}_n$ the average of $\bar{S}_n$. The average percentage of people receiving sick pay *across* neighborhoods

[10] A LPM is an ordinary least squares (OLS) regression with a binary dependent variable.

in the sample $\overline{\overline{S}}_n$ is 20.97%, while the standard deviation of $\overline{S}_n$ is 5.54%.[11] In the following, I will try to explain this variation.

I started with the following LPM to examine if the local variation simply reflects observable socioeconomic factors:

$$S_{in} = \alpha + \beta X_{in} + \varepsilon_{in}, \qquad (1)$$

where *X* represents three types of socioeconomic variables: individual characteristics, characteristics of the individual's workplace, and neighborhood characteristics, as described in Appendix 1.[12] The socioeconomic variables explain very little of the local variation in sickness absence, with $R^2 = 0.0429$. Table 2 shows the estimated model.

To examine how much of the variation in the percentage of people on sick pay across neighborhoods that can be explained by *X*, I used the following generalized linear model (GLM):

$$\overline{S}_n = \alpha + \beta \overline{X}_n + \varepsilon_n. \qquad (1')$$

According to Papke and Wooldridge (1996) the GLM method is better than running a linear model on the logit-transformed dependent variable. It takes into account that one cannot do logit transformation of zeroes and ones, which would lead to missing values. Moreover GLM takes into account that the error terms have non constant variance (heteroscedasticity).

To estimate this model, the measures of the socioeconomic variables represented by *X* were changed to average neighborhood levels, reducing the number of observations from 1,527,533 to 13,123. The socioeconomic variables have relatively high explanatory power on the variation of average percentage of people on sick pay across neighborhoods. While the standard deviation of $\overline{S}_n$ for this "sample" is 8.61%, the standard deviation of the estimated dependent variable $\hat{\overline{S}}_n$ has dropped to 3.23% after controlling for the average socioeconomic

[11] The percentage of people receiving sick pay $\overline{S}_{in}$ in the 2003 sample as a whole is 20.98%.

[12] Note that the municipality specific variables listed in Appendix 1 are not included in *X*. I have not included income. The reason is that reported income is affected by the individual's sickness absence. Including income among the explanatory variables would have given rise to a bias in the estimates. Several of my explanatory variables are, however, correlated with income – for instance, age, education, and industry.

variables represented by $\overline{X}$. Nevertheless, there is still variation in sickness absence across neighborhoods left to be explained. See Table 2 for the model results.

Next I estimated a mixed-effect model to see if the remaining variation (the average residuals $\overline{\varepsilon}_n$), are systematic rather than random. Specifically, I included random neighborhood-specific intercepts μ_n :

$$S_{in} = \alpha + \beta X_{in} + \mu_n + \varepsilon_{in}. \tag{1''}$$

The reported likelihood-ratio test confirms that this random-intercept model offers significant improvement over LPM (1) with fixed effects only (p = 0.0000), indicating that there are systematic neighborhood effects. Table 2 shows the estimated model.

6.2. Measuring the Effect of Social Interactions in Neighborhoods

As discussed earlier, the main aim of the thesis is to investigate if local variations across neighborhoods reflect "group" effects on individual's probability of sickness absence. I adopt the well-used approach to examine group effects by including the percentage of individuals that received sick pay in a neighborhood[13], and estimate its effect in the following model: [14]

$$S_{in} = \alpha + \beta X_{in} + \gamma \overline{S}_n + \varepsilon_{in}. \tag{2}$$

As mentioned in section three, there are several methodological challenges related to the estimation of group effects using this approach. When I nevertheless ran the LPM (2), I obtained the estimate $\hat{\gamma}$ = 0.3996, which is significant at the one-percent level, as seen in Table 2.[15] The model has marginally higher explanatory power than LPM (1) (adjusted R^2 = 0.0457), suggesting presence of neighborhood effects.

6.2.1. Self-categorization and Social Influence

Festinger's (1954) social comparison theory postulates that humans have a drive to evaluate their opinions and abilities. The center of his theory is the "similarity hypothesis", which

[13] All percentage variables included in the models in this thesis are actually proportions, because these variables are not multiplied by 100.

[14] Individual i's is excluded from the neighborhood percentage. The variables for average age for each neighborhood and workplace also exclude individuals i's effect on the averages.

[15] To see if this high estimate was mostly affected by the small neighborhoods in my data, I ran this same model for those neighborhoods that has at least 10 inhabitants. I was able to reject this hypothesis.

predicts that individuals prefer to compare themselves with similar others. According to Turner (1985) it is those who are regarded as members of the same category or group as oneself who exert influence. Hence, social influence results from a process of self-categorization whereby the person perceives him- or herself as a group member possessing the same characteristics as other group members (Abrams, Wetherell, Cochrane, Hogg, & Turner, 1990). Abrams et al. (1990) conducted three experiments relating self-categorization and social influence. The subjects for the two first experiments were undergraduates who were enrolled in introductory psychology at the University of Dundee and participated as a course requirement. Subjects in the third experiment were high school student in the ages 16-17. The results from the three experiments suggested that self-categorization can be a crucial determining factor in social influence, and that the extent of normative influence may largely depend on whether the source of influence is regarded as a member of one's category (Abrams et al., 1990).

If we believe in the estimate from LPM (2) and give our trust in these theories and findings, we would expect to see a higher estimate for the parameter of interest if we were to run this model for more similar people. The hypothesis being that more similar people have higher influential power on each other. When I ran LPM (2) separate for women, the results were as predicted: the estimated coefficient increased to $\hat{\gamma}_F$ = 0.4497, which is significant at the one-percent level. However, I can't exclude that this higher estimate is merely a result of homophily, rather than a result of higher influence. To see if similar patterns exists for males, I also ran model (2) separate for men. The result showed just the opposite of what I found for women: the estimated coefficient decreased to $\hat{\gamma}_M$ = 0.3451.

The models I have looked at thus far give an indication that social interaction effects are present among neighbors. In the following section I will try to quantify group effects by adopting an approach used by Lindbeck et al. (2007) for dealing with the reflection problem.

6.3. Immigrants

To examine whether immigrants are affected by the absence behavior in the neighborhood where they settle down after arriving Norway, I used the following model:

$$S_{in}^{im} = \alpha + \beta X_{in}^{im} + \eta \overline{S}_{n}^{na} + \varepsilon_{in}. \qquad (3)$$

Here, S_{in}^{im} is the binary variable measuring if immigrant *i* living in neighborhood *n* received sick pay during the year, and $\overline{S}_{n}^{na}$ is the percentage of native Norwegians that received sick pay in that neighborhood. Since the absence variable on the left-hand side refers to a different group of people than the absence variable on the right-hand side, there is no reflection problem in this case (Lindbeck et al., 2007). The identifying assumption is that there is no tendency among immigrants with a high propensity for sickness absence to settle down in neighborhoods where the absence rates among natives are particularly high ("reverse causation") (Lindbeck et al., 2007).

A limitation that I face with this model is that I don't know when these "immigrants" moved to Norway. I only know that they came from abroad the last time they moved. There is thus a possibility that some "immigrants" have actually lived in the neighborhood longer than some of the "natives".

Since I want to study the influence of natives on immigrants, the analysis is confined to neighborhoods where the fraction of immigrants is less than 30 percent of total inhabitants.[16] The number of observations was drastically reduced to 37,673 as a result of missing data, the restriction I made, and the natural fact that most people living in Norway are not immigrants.

As shown in Table 2, I obtained the estimate $\hat{\eta} = 0.2355$ which is significant at one percent level. There is however a possibility that this result depend on selection rather than on social interaction, thereby violating the basic assumption (Lindbeck et al., 2007). Like Lindbeck et al. (2007) I cannot rule out the possibility for some indirect mechanism by which immigrants with a strong propensity to call in sick, by self-selection would wind up in neighborhoods with many Norwegians having the same propensity. If this is the case, this may create a selection bias in the regression. While Lindbeck et al. (2007) were able to address this problem by focusing on recent immigrants, I'm not able to do this because of the limitation of my data. They found that the longer an individual has been in Sweden, the higher is the estimated social interaction effect. This could be interpreted in at least two ways: 1) immigrants take time to observe and adjust their behavior to the behavior of native Swedes,

[16] To see if the estimate was mostly affected by neighborhoods with larger proportions of immigrants, I ran the same model for those neighborhoods that have less than, or equal to 20 and 10 percent immigrants. I was able to reject this hypothesis.

and 2) the longer an immigrant has been in Sweden the more likely is self-selection bias (Lindbeck et al., 2007).

6.4. Logit

Most of the models this far were estimated using OLS. As mentioned, there are several potentials challenges with this approach. Unless the probability of receiving sick pay is the same for all individuals, the variances of the error terms will not be the same across cases, leading to heteroscedasticity (Aldrich & Nelson, 1984).[17] Another potential problem arises from the fact that residuals are only free to take on two possible values; therefore they can't be normally distributed (Aldrich & Nelson, 1984). These two problems suggest that the estimated standard errors will be wrong. Also, the predicted value of the dependent variable is free to vary between negative infinity and positive infinity, whereas probabilities only can range between zero and one. Thus the OLS assumption of linearity is highly unreasonable when dealing with a binary dependent variable (Aldrich & Nelson, 1984).

Breusch-Pagan (BP) tests indicate that there is a significant degree of heteroscedasticity, and kernel density estimations (KDE) confirms that the residuals are not normally distributed. In the following I report the findings from the estimated logit regressions of the former estimated LPM's. In addition to the estimated coefficients of interest, I also report mean marginal effects implied by the logit models and give interpretations of these.[18]

6.4.1. Logit Model (1)

A likelihood ratio (LR) test shows that the model fits the data better than an intercept-only model (P = 0.0000). The linear predicted value from a "link test" in the statistical package Stata is statistically significant, indicating that the model is not completely misspecified. However since the link test is also significant, this usually means that I have either omitted relevant variable(s), or that the link function is not correctly specified. In addition, a Hosmer and Lemeshow's goodness-of-fit test indicates that the model fits the data poorly. Table 3 gives the estimated logit model (1).

[17] All linear probability models were also estimated using heteroscedasticity consistent or "White" standard errors. The statistical significance of the estimates and the estimated coefficients of these regressions did not differ from the standard OLS estimates.
[18] The mean marginal effects were obtained from post-estimation commands in Stata.

6.4.2. Logit Model (1'')

Judging both by the likelihood-ratio test comparing this model to an ordinary logit regression (*P*= 0.0000), and by the standard deviation of random intercepts (0.2004) being more than 62 times its standard error (0.0032), the random intercepts in the output exhibits significant variation. This model offers thus a significant improvement over logit model (1), as shown in Table 3.

6.4.3. Logit Model (2)

A LR test shows that the model fits the data better than model (1) (P = 0.0000). As seen from Table 3, I obtained the estimate $\hat{\gamma}$ = 2.4682, which is significant at the one-percent level. The mean marginal effect for this coefficient is 0.3899, also seen in Table 3. This mean marginal effect suggest that if the percentage of individuals on sick leave in one's neighborhood increase by one percent, the individual probability for sick leave increases with 0.39 percent.[19] As mentioned earlier however, I view this estimate mostly as an indication.

When I ran model (2) separate for men and women, the result was unexpected: the estimated effect for females decreased to $\hat{\gamma}_F$ = 2.3756, whereas the estimated effect for males increased to $\hat{\gamma}_M$ = 2.4947, both significant at the one-percent level. However, the mean marginal effect for these estimated coefficients were 0.4427 and 0.3275 respectively. Thus, the direction of these estimates is consistent with the prior results.

6.4.4. Logit Model (3)

As shown in Table 3, I obtained the estimate $\hat{\eta}$ = 1.5561 for logit model (3), which is significant at one percent level. The mean marginal effect (0.22196) suggest that if the percentage of Norwegians on sick leave in an immigrant's neighborhood increases with one percent, this immigrant's probability for sick leave increases with 0.24 percent. I consider this effect to be moderately large.

Thus far the aim has been to analyse social interaction affects among neighbors. Results from the estimated models are consistent, and indicate the presence of social interaction effects.

[19] Mean marginal effects show change in percentage points. However, since the average absence variable on the right-hand side only marginally differs from the absence variable on the left-hand side, the calculated change in percentages for this model is equal to the change in percentage points (when rounded to two digits). Any models where I report a percentage change that is equal to the mean marginal effect, rest on the same explanation.

Moreover, model (3) which seek to quantify the social interaction effect show a moderately large effect. However, for reasons mentioned in section three I can't establish endogenous social interaction effects as a central causal mechanism. In the next section I compare the models presented thus far over time. I then move on to study family effects before examining social interaction effects among colleagues. I finish my empirical analyse by furtherer examining neighborhood effects in a spatial context.

6.5. Comparing the Models across Time

Optimally I would want to estimate all the models using panel data. However, due to constraints, I had to do the analyses using cross-sectional data from 2003. Thus, I wanted to see if there were noteworthy differences between the years 2000-2003.The data and the variables are created the same way for all years. Table 4 reports some key comparisons of the models I have presented thus far. As seen from the table, although there is variation in the magnitude of the estimates, they are reassuringly close.

In terms of statistical significance there is only one noteworthy difference. The estimated coefficient $\hat{\eta}$ for model (3) is statistically significant only at a five percent level for the year 2000, compared to one-percent level for the remaining years. What's also interestingly is that the results from estimating model (2) separate for men and women, are the same across all years.[20] I'm thus able to conclude that the findings I have reported thus far are not restricted to the year 2003.

6.6. Measuring the Effect of Social Interactions in Families

To see if I could find any indications of social interaction effects among family members, I used the following model:

$$S_{if} = \alpha + \beta X_{if} + \chi \overline{S}_f + \varepsilon_{if} , \qquad (4)$$

where S_{if} is the binary variable measuring if individual *i* belonging to family *f* received sick pay in 2003, and $\overline{S}_f$ is the percentage of individual *i*'s family on sick leave. Besides this, the model specification is identical to model (2).[21] As seen from Table 5, I obtained the estimates

[20] Although not reported in Table 4, mean marginal effects were estimated all years to confirm this.
[21] Individual *i* was excluded from the absence rate.

$\hat{\chi}$ = 0.0730, and $\hat{\chi}$ = 0.4636 for LPM (4), and logit model (4) respectively. Both estimates are significant at the one-percent level. Also as seen in Table 5, the mean marginal effect is 0.0693, suggesting an increased probability for sick leave of 0.07 percent if the percentage on sick leave in one's family increases with one percent.

There is thus indication of social interaction effects among family members. It's worth mentioning that the estimated percentage change is smaller than the one estimated for logit model (2). Also note that the number of observations for this model is smaller than for model (2), because those individuals without family members were excluded from the regression.

6.7. Measuring the Effect of Social Interactions at Workplace

Next I wanted to analyse social interaction effects on sickness absence at the workplace. Here I used model (1) from Dale-Olsen et al.'s study (2010) as a benchmark. My model specification is as followed:

$$S_{iw} = \alpha_1 + \alpha_2 Z_{iw} + \alpha_3 \overline{Z}_{jw} + \alpha_4 \overline{S}_w + \varepsilon_{iw}, \tag{5}$$

where S_{iw} is the binary variable measuring if individual *i* working at workplace *w* received sick pay in 2003, Z_{iw} represents three types of socioeconomic variables; individual characteristics, characteristics of the individual's workplace, and characteristics of the municipality where the workplace is located[22], as described in Appendix 1.[23] $\overline{Z}_{jw}$ represents the individual specific characteristics reported in Appendix 1 of individual *i*'s average colleague *j*. $\overline{S}_w$ is a measure of the percentage of *i*'s colleagues on sick pay during the year.[24]

As shown in Table 6, for LPM (5) I obtained the estimate $\hat{\alpha}_4$ = 0.2968 which is significant at the one-percent level. Running logit model (5) I obtained the estimate $\hat{\alpha}_4$ = 1.8139, also significant at a one-percent level. The mean marginal effect is 0.2848. If the percentage of one's colleagues on sick leave increases with one percent, it suggests an increased probability of 0.28 percent for receiving sick pay. Thus, I find indications of social interaction effects among colleagues. Note that the number of observations for this model is smaller than for the previous models because of missing data on where the workplace is located.

[22] I don't have data for in which basic statistical unit the workplace is located, only the municipality.
[23] Note that the neighborhood specific variables listed in Appendix 1 are not included in *Z*.
[24] Individual *i* was excluded from the absence rate.

6.8. Spatial Analysis

Thus far the focus has been on how individuals belonging *in* groups affect each other's sickness absence behavior. In the following I introduce spatial econometrics to more closely examine neighborhood effects. Specifically I look at how nearby geographical areas affect the sickness absence of inhabitants in a focal neighborhood.

6.8.1 Operationalization of the Weight Matrix

The influence structure in a network is represented by a weight matrix, where each row displays the influence on an actor and the column displays the influence exerted by this actor (Leenders, 2002). With regard to operationalization of the weight matrix, two components play a role: the choice for an operationalization of nearest, and the choice for a particular normalization (Leenders, 2002). The former defines who constitutes the actor's frame of reference, and the latter determines how social influence is allocated among these influencers (Leenders, 2002).

In this thesis the bordering basic statistical units of neighborhood *n* (the first-order neighbors) are defined as the influencers, and those that do not border the neighborhood are assumed to have no influence. The weight matrix constructed is row normalized. This entails that if neighborhood *n* has three first-order neighbors, each of these influencers will have weight 1/3. I therefore assume that all first-order neighbors exert the same amount of influence on that particular neighborhood. A neighborhood with only one first-order neighbor will be fully influenced by this one neighbor (weight one).

6.8.2. Model Specification

Spatial regression models assume that individuals, or more generally units of analysis (in this case basic statistical units) can be located in a space (Bradlow et al., 2005). Typically, actions by individuals are assumed to be correlated in such a manner that individuals near one another in space generate similar outcomes (Bradlow et al., 2005). Using ordinary least squares to analyse this type of sample data has been found to produce residuals that vary systematically over space, a phenomenon known as spatial autocorrelation (LeSage, 2000).

The spatial model is specified to include spatial lags of the dependent variable, of the explanatory variables, or spatial lags to reflect dependence in the disturbance process. Inclusion of one type of spatial lag does not mutually exclude another, and it is possible to include all three types of spatial lag in one model. In the context of sickness absence, I assume that spatial spillovers emerge through the dependent variable. In other words, I assume that one's sickness absence directly depend on the sickness absence in one's neighboring neighborhoods. Thus, I include spatial lag of the dependent variable in the spatial model, which gives me the spatial simultaneous autoregressive (SAR) lag model.

6.8.3. Measuring Spatial Effects at Neighborhood Level

To examine how first-order neighbors affect the average sickness absence in neighborhood *n*, I ran a maximum likelihood estimation of the following SAR lag model:[25]

$$\overline{S}_n^{\log} = \alpha + \beta \overline{X}_n + \rho W \overline{S}_{ne} + \varepsilon_n , \tag{6}$$

where *W* is the weight matrix, and *ne* denotes first-order neighbors. The dependent variable is here logit-transformed, hence the subscription *log*. [26] The mean continuous age variables were here scaled to fit the magnitude of the other variables.[27] All other variables are the same as in model (1'). Because I didn't have available information about the first-order-neighbors for some neighborhoods, these neighborhoods were excluded from the regression. The number of observations is thus smaller than for model (1').

As seen from Table 7, I was able to find a positive and highly significant neighboring neighborhood effect ($\hat{\rho}$ = 0.05348) which reflects the clear spatial pattern that is exhibited in Figure 1.[28] The odds ratio for this estimate is 1.0549[29], and the corresponding percentage change in odds ratio is 5.49. This shows that the odds will increase with 5.49% if the percentage of first-order neighbors on sick leave increases with one percentage point. For this model specification odds represent the percentage of individuals on sick leave, divided by the percentage of individuals who are not on sick leave. Thus, there is indication that the percentage of individuals on sick leave in a focal neighborhood is affected by the percentage sickness absence in first-order neighborhoods.

[25] Models (1) to model (5) were estimated using the statistical package Stata. The spatial analysis was however estimated using the statistical package R.

[26] Prior to the logit-transformation, I manipulated observations that took the values zero or one, for the reasons mentioned in section 6.1. The observations were changed to 0.01 and 0.99 respectively.

[27] Mean age, mean age squared, and mean age quadratic, were divided by 100, 1,000, and 10,000 respectively.

[28] This map was created based on the 1,527,533 individuals living in 13,123 different neighborhoods in the 2003 sample.

[29] The odds ratio is calculated as $e^{0.05348}$, and the percentage change in odds ratio is calculated as $[(e^{0.05348*100}) - (e^{0.05348*99}) / (e^{0.05348*99})]*100$. This change is constant for all adjacent pairs of values.

7. Conclusions and Discussions

In this thesis I investigated whether social influence play a role in explaining observed patterns in sickness absence in Norway. Although I examine social interaction effects among family members, colleagues and neighbors, the main focus is on the latter. The results show a consistent indication of social interaction effects for all three groups.

Due to the methodological challenges discussed in this thesis, one should be cautious before claiming the presence of endogenous social interaction effects. Logit model (3), where the reflection-problem is controlled for, is one attempt to more carefully assess social interaction effects among neighbors. The estimated effect suggest that if the percentage of Norwegians on sick leave in a neighborhood increase with one percent, an immigrant's probability for sick leave (above 16 days) increases with 0.24 percent. In addition, the spatial analysis conducted at the neighborhood level indicates that the percentage of individuals on sick leave in nearby geographical areas affects sickness absence in a focal neighborhood.

In addition to the methodological challenges discusses in this thesis, it is worth mentioning a key assumption made about the groups that I study. This assumption is that group members (as defined in this thesis) have social contact. For families this is likely to be a fairly realistic assumption. However for colleagues and neighbors the assumption is fairly strict. That being said, a particular strength of the data set used in this thesis is the detailed geographical data. Compared to using municipality (e.g., Bokenblom & Ekblad, 2007) which are far coarser geographical regions, basic statistical units in Norway were subdivided to give general and stable geographical units more suitable for statistics within smaller geographic areas (Byfuglien & Langen, 1983). One could argue that a natural assumption would be that the finer the division into "neighborhoods", the more likely one is to capture those neighbors who interact. On the other hand, Christakis and Fowler (2012) have found evidence form different health behavior studies indicating that a friend who lives hundreds of miles away appears to have a similar effect as a friend who live next door, and thus that social distance appears to matter much more than physical distance. Not knowing the exact relationship other than assuming that there is some sort of interaction between colleagues, neighbors, and family members, this highlights the importance of network data to identify endogenous social interaction effects. In recent years there have been real progress in obtaining unbiased estimates of group effects using field experiments over social networks (e.g., Aral & Walker, 2011a; Castillo & Carter, 2003; Nickerson, 2008). A promising strategy for understanding the dynamics of social influence at scale are randomized trials (Aral & Walker, 2011b).

The analysis in this thesis was conducted on cross-sectional data. As we know, there are however many advantages by applying panel data, one of them being able to control for unobserved heterogeneity. The spatial analysis in this thesis was only conducted at neighborhood level, as a result of computationally challenges. Besides the needs for network data and field experiments, applying panel data, and bringing the spatial analysis down to individual level, could provide new and interesting insights in explaining observed patterns in sickness absence.

References

Abrams, Dominic, Wetherell, Margaret, Cochrane, Sandra, Hogg, Michael A, & Turner, John C. (1990). Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization*. *British Journal of Social Psychology, 29*(2), 97-119.

Aldrich, John H., & Nelson, Forrest D. (1984). *Linear probability, logit, and probit models*. Beverly Hills & London: Sage Publication.

Arai, Mahmood, & Thoursie, Peter Skogman. (2004). Sickness absence: Worker and establishment effects. *Swedish Economic Policy Review, 11*, 9-28.

Arai, Mahmood, & Thoursie, Peter Skogman. (2005). Incentives and selection in cyclical absenteeism. *Labour Economics, 12*(2), 269-280.

Aral, Sinan, Muchnik, Lev, & Sundararajan, Arun. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences, 106*(51), 21544-21549.

Aral, Sinan, & Walker, Dylan. (2011a). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science, 57*(9), 1623-1639.

Aral, Sinan, & Walker, Dylan. (2011b). Identifying social influence in networks using randomized experiments. *Intelligent Systems, IEEE, 26*(5), 91-96.

Bokenblom, Mattias, & Ekblad, Kristin. (2007). Sickness Absence and Peer Effects-Evidence from a Swedish Municipality: Working Paper.

Bradlow, Eric T, Bronnenberg, Bart, Russell, Gary J, Arora, Neeraj, Bell, David R, Duvvuri, Sri Devi, . . . Yang, Sha. (2005). Spatial models in marketing. *Marketing Letters, 16*(3-4), 267-278.

Bramoullé, Yann, Djebbari, Habiba, & Fortin, Bernard. (2009). Identification of peer effects through social networks. *Journal of econometrics, 150*(1), 41-55.

Byfuglien, Jan, & Langen, Ole Ragnar. (1983). Grunnkretser, tettsteder og menigheter: Dokumentasjon 1980, Rapport 13/83. Statistisk Sentralbyrå.

Castillo, Marco, & Carter, Michael. (2003). Identifying social effects with economic field experiments. Department of Agricultural and Applied Economics, University of Wisconsin, Working Paper.

Chadwick-Jones, John K, Nicholson, Nigel, & Brown, Colin. (1982). *Social psychology of absenteeism*: Praeger New York.

Christakis, Nicholas A, & Fowler, James H. (2012). Social contagion theory: examining dynamic social networks and human behavior. *Statistics in Medicine*.

Dale-Olsen, Harald, Nilsen, Kjersti Misje, & Schøne, Pål. (2010). Imitation, Contagion, and Exertion–Do Colleagues' Sickness Absences Influence your Absence Behaviour? Oslo, Norway: ISF Working paper 2010:010.

Duflo, Esther, & Saez, Emmanuel. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics, 118*(3), 815-842.

Festinger, Leon. (1954). A theory of social comparison processes. *Human relations, 7*(2), 117-140.

Folketrygdloven. (1997). Lov om folketrygd from http://www.lovdata.no/all/hl-19970228-019.html

Grodner, Andrew, & Kniesner, Thomas J. (2006). Social interactions in labor supply. *Journal of the European Economic Association, 4*(6), 1226-1248.

Holm, Anders, Jæger, Mads Meier, & Pedersen, Morten. (2008). Unobserved Heterogeneity in the Binary Logit Model with Cross-Sectional Data and Short Panels: A Finite Mixture Approach. University of Copenhagen. Centre for Applied Microeconometrics: Working paper no. 2009-04

Johansson, Per, & Palme, Mårten. (1996). Do economic incentives affect work absence? Empirical evidence using Swedish micro data. *Journal of Public Economics, 59*(2), 195-218.

Kossinets, Gueorgi, & Watts, Duncan J. (2009). Origins of homophily in an evolving social network1. *American Journal of Sociology, 115*(2), 405-450.

Lalive, Rafael, & Cattaneo, M Alejandra. (2009). Social interactions and schooling decisions. *The Review of Economics and Statistics, 91*(3), 457-477.

Leenders, Roger Th AJ. (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks, 24*(1), 21-47.

LeSage, James P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis, 32*(1), 19-35.

Lindbeck, Assar, Palme, Marten, & Persson, Mats. (2007). Social interaction and sickness absence. *IFN Working Paper No 725*.

Manski, Charles F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies, 60*(3), 531-542.

Manski, Charles F. (2003). Identification problems in the social sciences and everyday life. *Southern Economic Journal, 70*(1), 11-21.

Markussen, Simen, & Røed, Knut. (2012). Social insurance networks. IZA Discussion paper No. 6446.

McPherson, Miller, Smith-Lovin, Lynn, & Cook, James M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology, 27*, 415-444.

Murray, Michael P. (2006). *Econometrics: A modern introduction*: Pearson Addison Wesley.

Nickerson, David W. (2008). Is voting contagious? Evidence from two field experiments. *American Political Science Review, 102*(1), 49-57.

NOU 2009: 10. (2009). *Fordelingsutvalget*. Oslo: Departementenes servicesenter. Retrieved from http://www.regjeringen.no/nb/dep/fin/dok/nouer/2009/nou-2009-10.html?id=558836.

NOU 2010: 1. (2010). *Medvirkning og medbestemmelse i arbeidslivet*. Oslo: Departementenes servicesenter. Retrieved from http://www.regjeringen.no/nb/dep/ad/dok/nouer/2010/NOU-2010-01.html?id=597723.

NOU 2011:1. (2011). *Bedre rustet mot finanskriser*. Oslo: Departementenes servicesenter. Retrieved from http://www.regjeringen.no/nb/dep/fin/dok/nouer/2011/nou-2011-1.html?id=631151.

OECD. (2012). OECD Economic Surveys: Norway 2012, OECD Publishing. doi:10.1787/eco_surveys-nor-2012-en

Papke, Leslie E, & Wooldridge, Jeffrey. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics, 11*, 619-632.

Rege, Mari, Telle, Kjetil, & Votruba, Mark. (2012). Social Interaction Effects in Disability Pension Participation: Evidence from Plant Downsizing*. *The Scandinavian Journal of Economics, 114*(4), 1208-1239.

Rieck, Karsten , Marshall, Elseth, & Vaage, Kjell. (2012). SOCIAL INTERACTIONS AT THE WORKPLACE: EXPLORING SICKNESS ABSENCE BEHAVIOR. *Memo, Dept. of Economics, University of Bergen*.

Statistisk Sentralbyrå. (1999). Regionale inndelinger-En oversikt over standarder i norsk offisiell statistikk. Retrieved March 24, 2013, from http://www.ssb.no/a/publikasjoner/pdf/nos_c513/nos_c513.pdf

Statistisk Sentralbyrå. (2013a). Deltidsarbeid - blir forskjellene utjevnet? *Samfunnsspeilet, 2/2013*. Retrieved May 07, 2013, from http://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/deltidsarbeid-blir-forskjellene-utjevnet

Statistisk Sentralbyrå. (2013b). Sykefravær, 4. kvartal 2012. Retrieved May 02, 2013, from http://www.ssb.no/arbeid-og-lonn/statistikker/sykefratot/kvartal/2013-03-14#content

Turner, John C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. *Advances in group processes: Theory and research*, *2*, 77-122.

Table 1. Summary statistics of socioeconomic variables

	Obs.	Mean	Std.Dev.	Min	Max
Age	1,527,553	40.1330	11.6846	18	64
Age squared	1,527,553	1,747.19	952.6874	324	4,096
Age quadratic	1,527,553	81,131.67	62,608.09	5,832	262,144
Education level 1	1,527,553	0.0295	0.1693	0	1
Education level 2	1,527,553	0.3094	0.4622	0	1
Education level 3	1,527,553	0.4238	0.4942	0	1
Education level 4	1,527,553	0.2373	0.4254	0	1
Female	1,527,553	0.4789	0.4996	0	1
Not married	1,527,553	0.3788	0.4851	0	1
Married	1,527,553	0.5018	0.4999	0	1
Divorced/separated	1,527,553	0.1090	0.3117	0	1
Widowed	1,527,553	0.0104	0.1016	0	1
Has children aged 11 or younger	1,527,533	0.3056	0.4607	0	1
Born in Norway	1,527,533	0.936	0.2448	0	1
Industry type 1	1,527,533	0.0073	0.0853	0	1
Industry type 2	1,527,533	0.0191	0.1371	0	1
Industry type 3	1,527,533	0.1371	0.3440	0	1
Industry type 4	1,527,533	0.0088	0.0934	0	1
Industry type 5	1,527,533	0.0630	0.243	0	1
Industry type 6	1,527,533	0.1399	0.3469	0	1
Industry type 7	1,527,533	0.0277	0.1642	0	1
Industry type 8	1,527,533	0.073	0.2601	0	1
Industry type 9	1,527,533	0.0256	0.1576	0	1
Industry type 10	1,527,533	0.0915	0.2883	0	1
Industry type 11	1,527,533	0.0808	0.2883	0	1
Industry type 12	1,527,533	0.099	0.2986	0	1
Industry type 13	1,527,533	0.1965	0.3973	0	1
Industry type 14	1,527,533	0.0307	0.1726	0	1
Number of employees 1	1,527,533	0.3376	0.4729	0	1
Number of employees 2	1,527,533	0.1728	0.3780	0	1
Number of employees 3	1,527,533	0.1532	0.3601	0	1
Number of employees 4	1,527,533	0.2225	0.416	0	1
Number of employees 5	1,527,533	0.1139	0.3178	0	1
Average age at workplace	1,527,533	40.1330	5.7768	18	62.5
Densely populated neighborhood	1,527,533	0.8036	0.3973	0	1
Average age in neighborhood	1,527,533	40.1332	2.3534	18	64
Average age in municipality	1,527,420	40.1333	0.9869	35.7	49.3

Table 2. Neighborhood effects: linear probability models 1 - 3

	Model (1)		Model (1")		Model (2)		Model (3)	
Adjusted R-squared	0.0429				0.0457		0.0535	
	Coefficients	Std. Err.	Coefficients	Std. Err.	Coefficients	Std. Err.	Coefficients	Std. Err.
Explanatory variables								
Age	0.07063***	0.00098	0.07009***	0.001	0.06984***	0.00098	0.07341***	0.00612
Age squared	-0.00168***	0.00003	-0.00166***	0.00003	-0.00166***	0.00003	-0.00175***	0.00016
Age quadratic	0.00001***	2.05e-07	0.00001***	2.06e-07	0.00001***	2.04e-07	0.00001***	1.31e-06
Education level 1	0.07721***	0.00208	0.07391***	0.00208	0.07344***	0.00208	0.06055***	0.00657
Education level 2	0.10141***	0.00099	0.09494***	0.00101	0.0946***	0.00099	0.10887***	0.00605
Education level 3	0.0452***	0.00091	0.04112***	0.00092	0.04106***	0.00091	0.04405***	0.00521
Female	0.08416***	0.00073	0.08516***	0.00073	0.08509***	0.00073	0.08320***	0.00432
Not married	-0.06710***	0.00331	-0.06711***	0.00331	-0.06689***	0.00331	-0.10938***	0.02507
Married	-0.06671***	0.00322	-0.06541***	0.00322	-0.06531***	0.00322	-0.08587***	0.02469
Divorced/separated	0.00301	0.00333	0.00223	0.00333	0.00218	0.00333	-0.03456	0.02524
Has children aged 11 or younger	0.0258***	0.00083	0.02467***	0.00084	0.02451***	0.00083	0.02713***	0.00471
Born in Norway	-0.02142***	0.00137	-0.02067***	0.00138	-0.02086***	0.00137	-0.04147***	0.00418
Industry type 1	0.03246***	0.00433	0.03114***	0.00434	0.03076***	0.00432	0.04112	0.03133
Industry type 2	0.01033***	0.00311	0.01614***	0.00314	0.01363***	0.00311	-0.03349*	0.0181
Industry type 3	0.05347***	0.00223	0.04986***	0.00224	0.04909***	0.00223	0.02720*	0.01534
Industry type 4	0.02693***	0.00400	0.02330***	0.00401	0.02376***	0.00399	-0.00278	0.03763
Industry type 5	0.07879***	0.00246	0.07431***	0.00246	0.07402***	0.00246	0.05534***	0.01805
Industry type 6	0.01729***	0.00228	0.01465***	0.00228	0.01454***	0.00228	-0.01013	0.01577
Industry type 7	0.03409***	0.00293	0.03134***	0.00293	0.03165***	0.00292	0.03454**	0.01697
Industry type 8	0.07209***	0.00237	0.06881***	0.00237	0.06866***	0.00237	0.04915***	0.0161
Industry type 10	0.01219***	0.00232	0.01161***	0.00232	0.0116***	0.00232	0.00671	0.01517
Industry type 11	0.02678***	0.00232	0.02305***	0.00233	0.02313***	0.00232	0.00647	0.01649
Industry type 12	0.05492***	0.00232	0.04994***	0.00232	0.05016***	0.00232	0.01269	0.01544
Industry type 13	0.09932***	0.00218	0.09379***	0.00218	0.09355***	0.00217	0.04952***	0.01487
Industry type 14	0.03745***	0.00276	0.03683***	0.00276	0.03665***	0.00275	0.02138	0.01749
Number of employees 1	0.00806***	0.0012	0.00430***	0.00121	0.00408***	0.0012	0.01292*	0.00677
Number of employees 2	0.02237***	0.00129	0.01930***	0.0013	0.01893***	0.00129	0.02716***	0.00734
Number of employees 3	0.02409***	0.0013	0.0214***	0.00131	0.02101***	0.0013	0.03282***	0.00727
Number of employees 4	0.0226***	0.0012	0.02062***	0.00120	0.02040***	0.0012	0.01933***	0.00647
Average age at workplace	-0.00147***	0.00007	-0.00161***	0.00007	-0.00161***	0.00007	-0.00227***	0.00043
Densely populated neighborhood	-0.01085***	0.00083	-0.00874***	0.00099	-0.00482***	0.00084	0.00114	0.00723
Average age in neighborhood	-0.00087***	0.00014	-0.00087***	0.0002	-0.00114***	0.00014	-0.00288***	0.00077
Random-effects Parameter			0.03187	0.00051				
Percentage on sick leave in neighborhood					0.39959***	0.00593		
Percentage of natives on sick leave							0.23550***	0.02932
Constant	-0.73601***	0.01454	-0.7146***	0.01558	-0.78896***	0.01454	-0.68250***	0.09036
N	1,527,553		1,527,553		1,527,553		37,673	

Note: The dependent variable for each model is the binary variable for if the individual has been on sick leave or not during the year.

* Significant at 10%, ** significant at 5%, *** significant at 1%

Table 3. Neighborhood effects: logit models 1 - 3

	Model (1)		Model (1")		Model (2)		Model (3)	
Pseudo R-squared	0.0423				0.0451		0.0582	
	Coefficients	Std. Err.	Coefficients	Std. Err.	Coefficients	Std. Err.	Coefficients	Std. Err.
Explanatory variables								
Age	0.49474***	0.00672	0.49479***	0.00678	0.49124***	0.00673	0.57168***	0.0473
Age squared	-0.01159***	0.00017	-0.0116***	0.00017	-0.01151***	0.00017	-0.01356***	0.0012
Age quadratic	0.00009***	1.35e-06	0.00009***	1.36e-06	0.00009***	1.35e-06	0.00011***	9.70e-06
Education level 1	0.47429***	0.01248	0.45589***	0.01259	0.45106***	0.01252	0.42861***	0.04558
Education level 2	0.6047***	0.00627	0.56854***	0.00639	0.56543***	0.00631	0.70054***	0.04098
Education level 3	0.28854***	0.00599	0.26589***	0.00607	0.26517***	0.00601	0.32352***	0.03807
Female	0.53943***	0.0047	0.54987***	0.00473	0.54731***	0.00471	0.59372***	0.03097
Not married	-0.31884***	0.01828	-0.32124***	0.01841	-0.31915***	0.01832	-0.59746***	0.14800
Married	-0.31776***	0.01765	-0.31166***	0.01776	-0.30975***	0.01768	-0.43413***	0.14474
Divorced/separated	0.05866***	0.01825	0.05437***	0.01837	0.05368***	0.01829	-0.13379	0.14798
Has children aged 11 or younger	0.15755***	0.00526	0.15129***	0.00532	0.15026***	0.00527	0.18227***	0.03254
Born in Norway	-0.13271***	0.00832	-0.1297***	0.00849	-0.12958***	0.00834	-0.2941***	0.02945
Industry type 1	0.23903***	0.02908	0.23104***	0.02930	0.22691***	0.02914	0.31808	0.21719
Industry type 2	0.05004**	0.02222	0.09103***	0.0225	0.07447***	0.02224	-0.45495***	0.16265
Industry type 3	0.38311***	0.01521	0.36320***	0.01535	0.35781***	0.01523	0.25536**	0.12173
Industry type 4	0.1911***	0.02753	0.17005***	0.02771	0.17264***	0.02757	-0.02917	0.31104
Industry type 5	0.55371***	0.01658	0.52906***	0.01669	0.52629***	0.01661	0.46802***	0.13656
Industry type 6	0.13233***	0.01562	0.11638***	0.01572	0.11627***	0.01565	-0.03977	0.12546
Industry type 7	0.25885***	0.0195	0.24210***	0.01963	0.24432***	0.01953	0.28706**	0.12999
Industry type 8	0.49572***	0.01589	0.47913***	0.01599	0.47715***	0.01592	0.41510***	0.12535
Industry type 10	0.08094***	0.01603	0.07823***	0.01612	0.07787***	0.01606	0.09221	0.12127
Industry type 11	0.18831***	0.01592	0.16694***	0.01602	0.16687***	0.01594	0.08054	0.12975
Industry type 12	0.37861***	0.01575	0.35073***	0.01586	0.3517***	0.01577	0.14051	0.12285
Industry type 13	0.60478***	0.01473	0.57436***	0.01485	0.57193***	0.01476	0.37115***	0.11802
Industry type 14	0.27445***	0.01846	0.27278***	0.01855	0.27086***	0.01849	0.21357	0.13478
Number of employees 1	0.04957***	0.00762	0.02684***	0.00777	0.02582***	0.00764	0.10204**	0.04883
Number of employees 2	0.13813***	0.00815	0.12029***	0.00827	0.11807***	0.00816	0.19912***	0.05231
Number of employees 3	0.14584***	0.00817	0.13001***	0.00828	0.12733***	0.00819	0.23304***	0.05154
Number of employees 4	0.13760***	0.00761	0.12531***	0.00770	0.12379***	0.00762	0.14444***	0.04736
Average age at workplace	-0.00881***	0.00045	-0.00973***	0.00045	-0.00962***	0.00045	-0.01505***	0.00303
Densely populated neighborhood	-0.06533***	0.00515	-0.05388***	0.00621	-0.02533***	0.00520	0.01035	0.04902
Average age in neighborhood	-0.00593***	0.0009	-0.00539***	0.00127	-0.00720***	0.00091	-0.02131***	0.00546
Random-effects Parameter			0.20047	0.0032				
Percentage on sick leave in neighborhood					2.46819***	0.03688		
					[0.38988]			
Percentage of natives on sick leave							1.55607***	0.19844
							[0.22196]	
Constant	-8.21780***	0.09862	-8.16429***	0.10533	-8.59250***	0.09905	-8.66151***	0.68681
N	1,527,553		1,527,553		1,527,553		37,673	

Note: The dependent variable for each model is the binary variable for if the individual has been on sick leave or not during the year. Mean marginal effects implied by logit models for the coefficient of interest in brackets. Reported pseudo R-squares are McFadden's R-squared.

* significant at 10%, significant at 5%, *** significant at 1%.

Table 4. Comparing the years 2000-2003

	Year 2000	Year 2001	Year 2002	Year 2003
Statistics etc.				
Number of individuals	1,360,714	1,553,955	1,548,562	1,527,533
Number of neighborhoods	13,136	13,149	13,142	13,123
Number of different industries	14	14	14	14
$\overline{S}_{in}$	19.39 %	19.96 %	20.53 %	20.98 %
$\overline{\overline{S}}_{n}$	19.38 %	19.96 %	20.53 %	20.97 %
Standard deviation of $\overline{S}_n$	5.48 %	5.37 %	5.49 %	5.54 %
R^2 LPM (1)	0.0424	0.0448	0.0437	0.0429
Standard deviation of $\overline{S}_n$ GLM (1')	9.72 %	9.35 %	9.35 %	8.61%
Standard deviation of $\hat{\overline{S}}_n$ in GLM (1')	3.26 %	3.25 %	3.13 %	3.23%
LR test mixed-effect model (1")	P = 0.0000	P = 0.0000	P = 0.0000	P = 0.0000
$\hat{\gamma}$ LPM (2)	0. 3683***	0.3921***	0.4056***	0.3996***
Adjusted R^2 LPM (2)	0.0449	0.0474	0.0466	0.0457
$\hat{\gamma}_F$ LPM (2)	0.4070***	0.4440***	0.4604***	0.4497***
$\hat{\gamma}_M$ LPM (2)	0.3261***	0.3405***	0.349***	0.3451***
$\hat{\eta}$ LPM (3)	0.1041**	0.2087***	0.3138***	0.2355***
LR test logit model (1)	P = 0.0000	P = 0.0000	P = 0.0000	P = 0.0000
LR test logit model (1")	P = 0.0000	P = 0.0000	P = 0.0000	P = 0.0000
$\hat{\gamma}$ logit model (2)	2.3848***	2.5095***	2.542***	2.4682***
LR test logit model (2)	P = 0.0000	P = 0.0000	P = 0.0000	P = 0.0000
$\hat{\gamma}_F$ logit model (2)	2.2445***	2.4289***	2.4738***	2.3756***
$\hat{\gamma}_M$ logit model (2)	2.4829***	2.5394***	2.5501***	2.4947***
$\hat{\eta}$ logit model (3)	0.7204**	1.4456***	2.0672***	1.5561***

* significant at 10%, ** significant at 5%, *** significant at 1%

Table 5. Family effects: linear probability & logit model 4

	Linear probability model (4)		Logit model (4)	
Adjusted R-squared/Pseudo R-squared	0.0449		0.0473	
	Coefficients	Std. Err.	Coefficients	Std. Err.
Explanatory variables				
Age	0.07633***	0.00141	0.60350***	0.01048
Age squared	-0.00185***	0.00004	-0.01438***	0.00026
Age quadratic	0.00001***	2.97e-07	0.00011***	2.11e-06
Education level 1	0.07094***	0.00315	0.43656***	0.01963
Education level 2	0.08287***	0.00138	0.50507***	0.00907
Education level 3	0.03328***	0.00125	0.21396***	0.00859
Female	0.088***	0.00102	0.60663***	0.00702
Not married	-0.07174***	0.00748	-0.35899***	0.0432
Married	-0.06072***	0.00735	-0.29456***	0.04217
Divorced/separated	0.00234	0.00777	0.04860	0.04464
Has children aged 11 or younger	0.02032***	0.00116	0.13293***	0.00796
Born in Norway	-0.02055***	0.00200	-0.12635***	0.01275
Industry type 1	0.02738***	0.00593	0.20329***	0.04284
Industry type 2	0.00716*	0.0043	0.02281	0.03238
Industry type 3	0.04461***	0.00302	0.33310***	0.02150
Industry type 4	0.02332***	0.00529	0.17005***	0.03830
Industry type 5	0.06590***	0.00333	0.49126***	0.02357
Industry type 6	0.01041***	0.00308	0.07807***	0.02209
Industry type 7	0.02631***	0.00407	0.21945***	0.02876
Industry type 8	0.05903***	0.00326	0.42576***	0.02281
Industry type 10	0.00418	0.00316	0.02386	0.02286
Industry type 11	0.02323***	0.00314	0.16811***	0.02242
Industry type 12	0.04538***	0.00311	0.31821***	0.02203
Industry type 13	0.08538***	0.00293	0.53329***	0.02071
Industry type 14	0.03146***	0.0038	0.24035***	0.02668
Number of employees 1	0.00698***	0.00167	0.04257***	0.01123
Number of employees 2	0.02085***	0.0018	0.13458***	0.01198
Number of employees 3	0.02034***	0.00182	0.12930***	0.01207
Number of employees 4	0.02129***	0.00169	0.13703***	0.01129
Average age at workplace	-0.00137***	0.0001	-0.00855***	0.00066
Densely populated neighborhood	-0.00852***	0.00113	-0.05411***	0.00745
Average age in neighborhood	-0.00304***	0.00022	-0.02128***	0.00146
Percentage on sick leave in family	0.07304***	0.00119	0.46364*** [0.06930]	0.00747
Constant	-0.71482***	0.02142	-9.03747***	0.15476
N	757,421		757,421	

Note: The dependent variable is the binary variable for if the individual has been on sick leave or not during the year. Mean marginal effect implied by logit model for the coefficient of interest in bracket. Reported pseudo R-squared is McFadden's R-squared.

* Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 6. Workplace effects: linear probability & logit model 5

	Linear probability model (5)		Logit model (5)	
Adjusted R-squared/Pseudo R-squared	0.0510		0.0505	
	Coefficients	Std. Err.	Coefficients	Std. Err.
Explanatory variables				
Age	0.07121***	0.00099	0.50000***	0.00683
Age squared	-0.00169***	0.00003	-0.01172***	0.00017
Age quadratic	0.00001***	2.06e-07	0.00009***	1.37e-06
Education level 1	0.06311***	0.00209	0.38949***	0.01271
Education level 2	0.08143***	0.00105	0.49006***	0.00665
Education level 3	0.03402***	0.00095	0.22644***	0.00626
Female	0.08418***	0.00079	0.54997***	0.00516
Not married	-0.06817***	0.0033	-0.32832***	0.01838
Married	-0.0649***	0.00321	-0.3082***	0.01773
Divorced/separated	0.00104	0.00332	0.04745***	0.01834
Has children aged 11 or younger	0.02463***	0.00083	0.15205***	0.00528
Born in Norway	-0.01617***	0.00139	-0.10157***	0.00858
Industry type 1	0.01002**	0.00435	0.08100***	0.02949
Industry type 2	0.00115	0.00316	-0.01874	0.02268
Industry type 3	0.02426***	0.00228	0.19028***	0.01562
Industry type 4	0.02014***	0.00402	0.14818***	0.0278
Industry type 5	0.04592***	0.00255	0.33863***	0.01724
Industry type 6	0.00016	0.00230	0.01751	0.01588
Industry type 7	0.00537*	0.00299	0.06302***	0.02002
Industry type 8	0.03525***	0.00241	0.25206***	0.01628
Industry type 10	0.00852***	0.00234	0.04062**	0.01627
Industry type 11	0.01902***	0.00234	0.13311***	0.01610
Industry type 12	0.05463***	0.00252	0.39564***	0.01719
Industry type 13	0.06311***	0.00227	0.38621***	0.01542
Industry type 14	0.02190***	0.00276	0.17021***	0.01865
Number of employees 1	-0.00318***	0.00124	-0.04421***	0.00797
Number of employees 2	0.00741***	0.00132	0.03094***	0.00841
Number of employees 3	0.00867***	0.00132	0.03708***	0.00839
Number of employees 4	0.01017***	0.00121	0.04723***	0.00773
Average age at workplace	-0.02159***	0.00316	-0.12067***	0.02081
Densely populated municipality	-0.00567***	0.00085	-0.03270***	0.00530
Average age in municipality	0.00428***	0.00036	0.02666***	0.00229
Average age squared of colleagues	0.00055***	0.00008	0.00311***	0.00055
Average age quadratic of colleagues	-4.53e-06***	7.13e-07	-0.00003***	4.67e-06
Percentage of colleagues with education level 1	0.01958**	0.00862	0.07734	0.05538
Percentage of colleagues with education level 2	0.01934***	0.00261	0.43058***	0.01676
Percentage of colleagues with education level 3	0.01934***	0.00262	0.15212***	0.01726
Percentage female colleagues	-0.02378***	0.00175	-0.15772***	0.01127
Percentage of colleagues not married	0.06240***	0.01523	0.48366***	0.09637
Percentage of colleagues married	0.01205	0.01497	0.13533	0.09466
Percentage of colleagues divorced/separated	0.07009***	0.01544	0.51914***	0.09751
Percentage of colleagues with children aged 11 or younger	0.00762**	0.00332	0.03710*	0.02143
Percentage of colleagues born in Norway	-0.04139***	0.00455	-0.27795***	0.02847
Percentage on sick leave at workplace	0.29681***	0.00319	1.81388***	0.01970
			[0.28483]	
Constant	-0.77245***	0.04383	-8.76656***	0.28945
N	1,527,420		1,527,420	

Note: The dependent variable is the binary variable for if the individual has been on sick leave or not during during the year. Mean marginal effect implied by logit model for the coefficient of interest in bracket. Reported pseudo R-squared is McFadden's R-squared. * Significant at 10%, ** significant at 5%, *** significant at 1%.

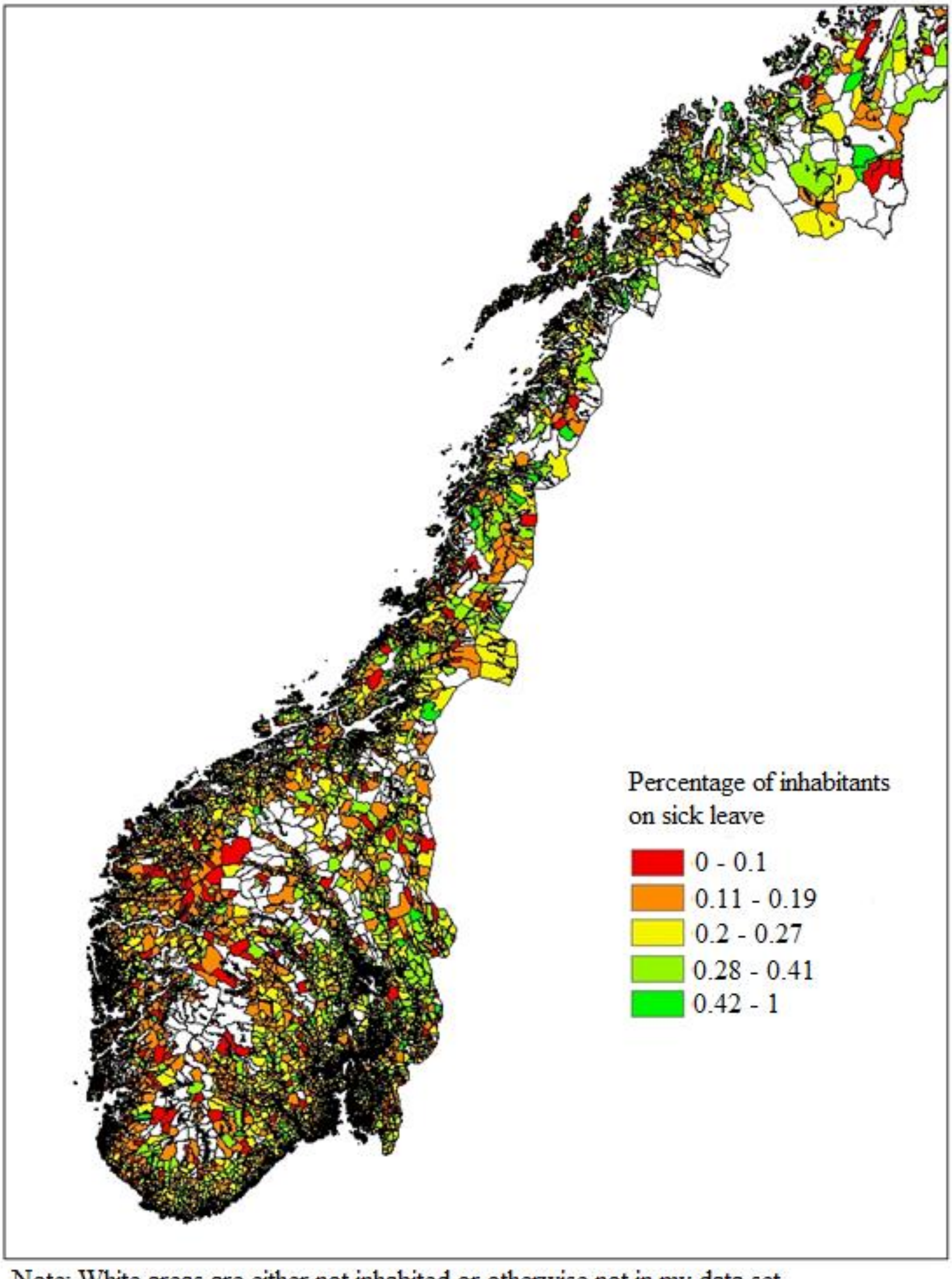
Table 7. Spatial effects at neighborhood level

	Spatial autoregressive model (6)	
Pseudo R-squared		
	Coefficients	Asymptotic Std. Err.
Explanatory variables		
Average age	58.78905***	7.45125
Average age squared	-14.31007***	1.90835
Average age quadratic	1.11638***	0.15574
Percentage with education level 1	1.19587***	0.18328
Percentage with education level 2	1.22862***	0.07671
Percentage with education level 3	0.44397***	0.08108
Percentage females	0.57678***	0.08032
Percentage not married	-0.86409***	0.10545
Percentage married	-0.91880***	0.08906
Percentage divorced/separated	0.14203	0.31090
Percentage with children aged 11 or younger	0.28713***	0.05925
Percentage born in Norway	0.30859***	0.11695
Percentage in industry type 1	-0.39284	0.28474
Percentage in industry type 2	-1.16451***	0.29019
Percentage in industry type 3	0.35025	0.25213
Percentage in industry type 4	-1.14204***	0.32639
Percentage in industry type 5	0.32811	0.26753
Percentage in industry type 6	0.08173	0.27047
Percentage in industry type 7	-0.54001*	0.28883
Percentage in industry type 8	-0.26017	0.26593
Percentage in industry type 10	-0.37186	0.28172
Percentage in industry type 11	0.11002	0.26079
Percentage in industry type 12	0.60292**	0.25799
Percentage in industry type 13	0.77463***	0.25026
Percentage in industry type 14	-0.34017	0.30389
Percentage with number of employees 1	-0.05064	0.08686
Percentage with number of employees 2	0.44316***	0.09791
Percentage with number of employees 3	0.45673***	0.10235
Percentage with number of employees 4	0.22786**	0.09874
Average of average age at workplace	-0.01241**	0.00551
Densely populated neighborhood	0.09355***	0.01438
Spatial neighborhood effect	0.05348***	0.01082
Constant	-9.38505***	0.98859
N	12,996	

Note: The dependent variable is the percentage of peoples in the neighborhood on sick leave. All explanatory variables are at average neighborhood level. Repported pseudo R-squared is Nagelkerke R-squared.

* Significant at 10%, ** significant at 5%, *** significant at 1%.

Figure 1. Map of percentage of individuals on sick leave in each neighborhood in Norway, 2003



Note: White areas are either not inhabited or otherwise not in my data set.

Appendix 1. Socioeconomic variables

For the individual	Age: ***Age*** (from 18 to 64, continuous) Age squared: ***Age squared*** (continuous) Age quadratic: ***Age quadric*** (continuous) Education: ***Education level 1, Education level 2, Education level 3, Education level 4*** (<9, 9-12, 13-15, ≥16; three dummies) Gender: ***Female*** (one dummy = 1 if female) Marital status: ***Not married, Married, Divorced/separated, Widowed*** (not married, married, divorced/separated, widowed; three dummies) Has children aged 11 or younger: ***Has children aged 11 or younger*** (one dummy = 1 if he/she has) Born in Norway: ***Born in Norway*** (one dummy = 1 if born in Norway)
For the workplace	Industry: ***Industry type 1, Industry type 2, … , Industry type 14*** (agriculture, mining, manufacturing, electronics, construction, wholesale, hotel, transport, financial, real estate, public administration, education, health, other; 13 dummies) Size of workplace: ***Number of employees 1, Number of employees 2, Number of employees 3, Number of employees 4, Number of employees 5*** (5-25, 25-50, 50-100, 100-500, > 500; 4 dummies) Average age at workplace: ***Average age at workplace*** (continuous)
For the neighborhood	Densely or sparsely populated area: ***Densely populated neighborhood*** (one dummy = 1 if densely) Average age in neighborhood: ***Average age in neighborhood*** (continuous)
For the municipality	Densely or sparsely populated area: ***Densely populated municipality*** (one dummy = 1 if densely) Average age in municipality: ***Average age in municipality*** (continuous)

Note: bold and cursive text shows the names used for the variables in the regression tables