IFN Working Paper No. 725, 2007

Social Interaction and Sickness Absence

Assar Lindbeck, Mårten Palme and Mats Persson

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Sickness8.doc

2007-12-13

Social Interaction and Sickness Absence*

Assar Lindbeck,[¤] Mårten Palme[#] and Mats Persson[♣]

Abstract
Does the average level of sickness absence in a neighborhood affect individual sickness absence through social interaction on the neighborhood level? To answer this question, we consider evidence of local benefit-dependency cultures. Well-known methodological problems in this type of analysis include avoiding the so-called reflection problem and disentangling the causal effects of group behavior on individual behavior from the effects of individual sorting on neighborhoods. Based on data from Sweden, we adopt several different approaches to deal with these problems. The results are robust in the sense that regardless of approach and identifying assumptions, we obtain statistically significant estimates indicating group effects.

JEL Codes H56, I38, J22, Z13

Keywords: Sick-pay insurance, work absence, moral hazard, social norms

---

[*] We are grateful for comments on a draft of the paper from seminar participants at the Economics Department and the Institute for International Economic Studies at Stockholm University, the Industrial Research Institute in Stockholm, the Economics Department at Uppsala University, and the 2007 meeting of the European Economic Association in Budapest.
[¤] Institute for International Economic Studies, Stockholm University, SE-106 91 Stockholm, Sweden and Research Institute of Industrial Economics, P.O. Box 55665, SE-102 15 Stockholm, Sweden. E-mail: Assar.Lindbeck@iies.su.se. Financial support from Sven and Catarina Hagströmer is gratefully acknowledged.
[#] Department of Economics, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: Marten.Palme@ne.su.se.
[♣] Institute for International Economic Studies, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: Mats.Persson@iies.su.se.

1. Introduction

The social insurance systems in advanced welfare states face serious moral-hazard problems. In some countries, these problems take the form of high levels of early retirement and disability pensions, such as in Belgium and the Netherlands. Other countries, like Norway and Sweden, have periodically experienced high and rapidly rising levels of sickness absenteeism – despite of indications of very good health conditions.[1] Sickness absenteeism has exhibited two conspicuous patterns: large aggregate cyclical fluctuations and huge variation across geographical areas within countries – even when the sick-pay insurance rules are the same throughout the country.

It has turned out to be difficult to explain the observed patterns by standard variables, such as benefit rules, socioeconomic factors and general (measurable) health conditions. Some observers have therefore argued that these patterns could be related to social interaction, such as group effects on individual behavior.[2] In this paper we study the importance of group effects on moral hazard in sick-pay insurance ("temporary disability insurance"), which is a major element of social insurance in Europe. Although such effects may encompass several different mechanisms, our hypothesis is that peer-group influence in the form of social norms (i.e., approval or disapproval by others) is one important mechanism.

We ask two questions. First, is there evidence that group influence exists in sickness absence behavior? Second, if such effects exist, how large might they be? These questions are important in the sense that group influence may accentuate the effects of exogenous changes, for instance in benefit rules, on sickness absence. Group influence thus may amplify the amount of moral hazard in insurance, an amplification that has been expressed by a so-called social multiplier in the theoretical literature.[3]

[1] Cf. Henrekson and Persson (2004) and Johansson and Palme (2005).
[2] For an attempt to document regional variations in attitudes concerning sickness absence, see Palmer (2006).
[3] See, e.g., Glaeser, Sacerdote and Scheinkman (2003) and Lindbeck and Persson (2006).

While group effects have been extensively analyzed theoretically,[4] empirical analysis has been held back by methodological problems. These problems stem from the obvious difficulty in distinguishing group effects on individual behaviour from other mechanisms which generate correlation between individual and group behaviour. Different methods to overcome this problem have been used in the empirical literature. One such attempt is to exploit exogenous variation in factors influencing group behavior (Hesselius and Johansson, 2005, and Duflo and Saez, 2003). Another is to study the consequences for individual behavior of simultaneously belonging to two groups with different behavior patterns, by introducing fixed group effects into the regression (Bertrand et al., 2000). A third approach has been to examine the importance of proximity among individuals for the transmission of behavioral patterns, either within a structural modelling framework (e.g. Glaeser, Sacerdote and Scheinkman, 2003) or by applying a reduced form approach (e.g. Bokenblom and Ekblad, 2007).

Using a data set that includes the entire Swedish population, we adopt four different approaches to analyze group effects on sickness absence behavior – with each approach requiring different identifying assumptions:

1. We exploit the difference in absence behavior between public- and private-sector employees to study whether the behavior of one of these groups of individuals influences the behavior of the other group.
2. We ask whether individuals who move from one neighborhood in Sweden to another tend to adjust their sick-absence behavior to normal behavior in the new neighborhood.
3. We study whether immigrants to Sweden adjust their behavior to that of native Swedes in the neighborhood where the immigrants have settled down.
4. We investigate the extent to which the behavior of an individual is influenced by the interaction of networks in his neighborhood and at his workplace.

[4] Examples are Moffit (1983), Lindbeck, Nyberg and Weibull (1999), Manski (2000) and Lindbeck and Persson (2000).

Under all four approaches, we define group effects as the influence of average behavior in a neighborhood on individual behavior. The first approach addresses the existence and the magnitude of such influence separately. The second approach aim is aimed at estimating the magnitude, while the third approach is confined to finding evidence for the existence of group effects.

As always, it is necessary to make rather specific identifying assumptions in this type of analysis, for instance concerning selection mechanisms. The advantage of using four different approaches (and alternative specifications within each) is that the study may then give an indication of the robustness of the results with respect to the underlying identifying assumptions.

2. A First Look at the Data

Our data set combines individual sickness absence data from the Swedish National Insurance Agency with a large number of socioeconomic variables obtained from the LISA database, compiled by Statistics Sweden. In addition to providing information on numerous individual characteristics, the combined data set allows us to identify each individual's neighborhood and workplace. The data consist of an unbalanced panel for the seven-year period 1996-2002. Although the data set covers the entire population in Sweden, we confine the study to private- and public-sector employees in the age group 18-64 (almost 5 million individuals, which implies about 25 million observations in the entire panel). A limitation in the data set is that it only covers spells of absence longer than 14 days.[5] It would have been of interest to study shorter spells as well, but such data are not available on an individual basis.

[5] The reason is that individual employers pay compensation for shorter spells, and that individual data on such spells are not systematically reported. The total average number of sick days for which sickness pay was claimed (including short spells) was about 25 per year during the period under study, as compared to 17.8 in our data set containing only absence spells longer than 14 days.

When studying local social norms, a first issue is to determine the most relevant geographical domain. Municipalities may be too large for this purpose. We have therefore chosen to use so-called Small Area for Market Statistics unit (SAMS) for geographical domains in Sweden.[6] Such areas provide reasonably homogeneous districts based on geographical proximity among inhabitants and similarity in housing.[7] There are 8,951 SAMS in our database, with an average population of 404 persons. In the following, we use the term "neighborhoods" for these areas.

It may be argued that social interaction takes place at both local and national levels. For example, mass media and the public policy debate can be important channels for social interactions on the national level. Similarly, local media and local organizations can be important for interaction, for instance on the county levels. Here, however, we focus on direct interaction on the personal level. For this purpose the SAMS seems to be an appropriate geographical unit.

We obtain a broad picture of local variations in sickness absence by looking at days of absence across neighborhoods during a year. For this purpose, we choose the last year for which we have data, namely 2002. Let S_{in} denote the number of sick days of individual i living in neighborhood n in 2002, and $\overline{S}_n$ the average number of sick days in that neighborhood. While the average number of sick days (above 14) in our data is 17.8, the standard deviation of $\overline{S}_n$ is 13.2 days per year. How can this wide variation across neighborhoods be explained?

First, to see whether the local variation simply reflects observable socioeconomic factors, we run a multivariate regression of the form

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

[6] See Statistics Sweden (2005) for a detailed description of this geographical specification.
[7] It turns out that our empirical results are approximately the same regardless of whether we use municipalities, church parishes or the SAMS as the basic geographical unit.

where the X vector contains three types of socioeconomic variables: individual characteristics (such as age and education), characteristics of the individual's workplace (such as industry and plant size), and neighborhood characteristics (such as urban/rural, local unemployment, and a local health variable). We have chosen explanatory variables that, in different studies, have turned out to be important for sickness absence. Due to the large number of observations, we can apply a flexible specification of the regression equation, using dummies rather than specific functional forms. A full list of the variables in the X vector is given in Appendix 1.[8]

As expected, the X vector explains very little of each individual's behavior, since idiosyncratic factors tend to dominate individual behavior. More surprisingly, the X vector also explains very little of the variation of *average* sickness absence, $\bar{S}_n$, across neighborhoods. While the standard deviation of average absence across neighborhoods in the 2002 raw data was 13.2 days, it is almost the same (12.9 days) after controlling for all the socioeconomic variables in the X vector. To find out whether the remaining differences among neighborhoods (the average residuals $\bar{\varepsilon}_n$) are systematic rather than random, we estimate an equation with neighborhood-specific intercepts α_n:

$$S_{in} = \alpha_n + X_{in}^{'}\beta + \varepsilon_{in} . \qquad (1')$$

An F test suggests that (1') fits the data significantly better than the original specification (1) with a uniform intercept ($F = 2.650$, implying significance at the one-percent level[9]). To rule out the possibility that this simply reflects permanent unobservable factors, we also estimate (1) and (1') in terms of changes in sickness absence. As in the case of levels,

---

[8] We have not included income in the X vector. 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 our explanatory variables are, however, correlated with income – for instance, age, education, gender, and industry.
There are arguments for and against including local unemployment among the explanatory variables. In this paper, we have chosen to report the results from regressions where local unemployment *is* included – although excluding it would not change the results noticeably in terms of the influence of social norms on individual sickness absence.
[9] See, for instance, Greene (2003, chapter 13).

the average residuals of changes between 2001 and 2002 across neighborhoods vary systematically, i.e., in a non-random fashion (F = 1.370, again implying significance at the one-percent level). Thus, there is systematic local variation in average sickness absence not accounted for by the socioeconomic factors in our X vector. This holds not only for levels, but also for changes. Indeed, this result holds for the entire panel, and not only for specific years.

3. Measuring the Effect of Social Interactions

The aim of our study is to investigate whether these large local variations reflect group effects on individual sickness absence behavior and, if so, how strong such effects are. We thus measure group behavior by the average number of sickness absence days in a neighborhood, by estimating the following relation:

$$S_{in} = \alpha + X_{in}'\beta + \gamma \overline{S}_n + \varepsilon_{in} \,. \qquad (2)$$

As is well known, there are several methodological problems related to the estimation of group effects.[10] One serious problem is how to separate out the effect of group behavior through social interaction from the effect due to the fact that individuals with similar unobserved characteristics tend to live in the same neighborhood (correlated effects) or be exposed to similar local differences in policy (contextual effects). Indeed, running an OLS regression on (2) tends to give a biased estimate of γ because of the so-called reflection problem (Manski, 2000): on average, an individual's behavior is tautologically related to the average individual's behavior. When we nevertheless run an OLS regression on (2), we obtain the estimate $\hat{\gamma} = 0.8658$, which is significant at the one-percent level.

[10] In all regression with average sickness $\overline{S}_n$ as an explanatory variable, we exclude the individual's own absence from the neighborhood average.

As described in the Introduction, we use four approaches for dealing with the reflection problem. One is based on exogenous variation in neighborhood sickness absence as the result of differences in absenteeism between employees in the public and private sector. The other approaches – the analysis of movers and immigrants, and of interactions between different networks – rely on controls for neighborhood fixed effects.[11]

4. Public-sector vs. Private-sector Employees

Public-sector employees in Sweden have systematically higher sickness absence than private-sector employees.[12] There may be several reasons behind this empirical regularity. The most obvious is that private employers have stronger incentives to prevent absence, since it is costly to the employer, whereas public employers have weaker direct incentives to minimize costs to their organization. It could also be the case that workers with preferences for frequent absence value the higher degree of employment security in the public sector and therefore self-select the public sector.

This, in turn, means that neighborhoods with a large share of public-sector employees are, on average, likely to have a higher work absence rate. We exploit this fact and use the share of public-sector employees as an instrumental variable for the average work-absence level in the neighborhood. We then carry out the analysis separately for private- and public-sector employees, respectively. The identifying assumption underlying this approach is that the share of public-sector employees in the neighborhood is unrelated to unobserved characteristics affecting individual work-absence behavior; formally, $E(Z_{nt}\varepsilon_{\text{int}}) = 0$, where Z_{nt} is the public sector's share of employment in neighborhood *n* in year *t*.[13] Thus we assume that workers with specific absence behavior do not choose to

[11] A model similar to that in equation (2) including individual fixed effects or differencing of the data would not solve the reflection problem, since the underlying factors (i.e., unobserved heterogeneity) may operate on changes as well as on levels.
[12] By sectors, the average number of days of sickness absence in our data set (spells longer than 14 days) in 2001 were: private-sector employees 12.2; central government employees 15.4; municipal employees 20.3.
[13] More exactly, *Z* is the ratio of the number of public-sector employees to the sum of public- and private-sector employees.

settle down in neighborhoods on the basis of the proportion between public- and private-sector employees in these neighborhoods. In other words, we assume that the different behavior of these two groups of employees is related to the institutional characteristics of the sectors where they work, rather than to unobserved individual differences.

We use the following IV model to explain the behavior of individual private-sector employees:

$$\begin{aligned} \overline{S}_{nt} &= a + k_t + X'_{\mathrm{int}}\beta + cZ_{nt} + e_{nt} \\ S^{priv}_{int} &= \alpha + \lambda_t + X'_{int}\beta + \gamma\hat{\overline{S}}_{nt} + \varepsilon_{int}. \end{aligned} \tag{3}$$

Conversely, we use $1 - Z_{nt}$ as an instrument to estimate the effects on individual public-sector employees, S^{publ}_{int} :

$$\begin{aligned} \overline{S}_{nt} &= a + k_t + X'_{\mathrm{int}}b + c(1 - Z_{nt}) + e_{nt} \\ S^{publ}_{int} &= \alpha + \lambda_t + X'_{int}\beta + \gamma\hat{\overline{S}}_{nt} + \varepsilon_{int}. \end{aligned} \tag{3'}$$

We also estimate a system like (3) and (3') for the entire population, i.e., without superscript *priv* or *publ* on the *S* in the second equation.

Before pursuing the IV analysis, it is of interest to investigate more directly whether the absence behavior of a private-sector employee is higher if he has many neighbors who work in the public sector, and vice versa. We therefore study the reduced form of the model defined in equations (3), i.e.,

$$S^{priv}_{int} = \alpha + \lambda_t + X'_{int}\beta + \mu \cdot Z_{nt} + \varepsilon_{int}\,, \tag{4}$$

where S^{priv}_{int} is the sickness absence of individual *i* working in the private sector in year *t*. We expect the estimate of μ to be positive.

Conversely, we ask whether a public-sector employee tends to be less absent from work if he lives in neighborhoods where there are many private-sector employees:

$$S_{int}^{publ} = \alpha + \lambda_t + X_{int}^{'}\beta + \mu \cdot (1 - Z_{nt}) + \varepsilon_{int} \,. \tag{4'}$$

The results of the estimates are shown in the fourth column of Table 1. As expected, a higher share of public-sector employees in a neighborhood is associated with higher sickness absence among private-sector employees in that neighborhood. The number 0.0387 means that if the share of public-sector employees is 10 percentage points higher in one neighborhood than in another, then sickness absence among the privately employed is approximately 0.387 days higher in the first neighborhood. Similarly, if the share of private employees in one neighborhood is 10 percentage points higher in than in another, the number of sick days among public-sector employees is 0.438 days lower.

For completeness, we have also made reduced-form estimates based on the entire population (private- as well as public-sector employees). The number 0.0418 means that if the share of public employees in a neighborhood is 10 percentage points higher than in another, the average number of absence days among all employees is 0.418 days higher.

Note that the estimation of μ does not provide a quantification of the influence of average behavior on the individuals; it is simply an *indicator* of the existence of social interaction across the groups of private-sector and public-sector employees. Moreover, this indicator reflects only social interaction *across* the groups, not social interaction *within* groups. To obtain a quantification of total group influence on individual behavior, an estimate of γ in the full IV model is required. The resulting estimates are shown in the fifth and sixth columns of Table 1.

Table 1: Estimates of μ in (4) and (4'), and of γ in (3) and (3')

Population	Number of individuals and observations	Regressor	Reduced form (μ in eq. (4))	First step in IV regression (c in eq. (3))	IV estimate (γ in eq. (3))
All those who work in private sector	2,839,410 ind. 14,556,753 obs.	Share of population in neighborhood n that work in public sector (Z_{nt})	0.0387*** (0.0013) $R^2 = 0.020$	6.670*** (0.0116) $R^2 = 0.499$	0.581*** (0.0199) $R^2 = 0.0215$
All those who work in public sector	1,956,740 ind. 10,502,405 obs.	Share of population in neighborhood n that work in private sector ($1 - Z_{nt}$)	-0.0438*** (0.0017) $R^2 = 002$	-5.752*** (0.0123) $R^2 = 0.512$	0.762*** (0.0302) $R^2 = 0.274$
All employees	4,796,150 ind. 25,059,158 obs.	Share of population in neighborhood n that work in public sector (Z_{nt})	0.0418*** (0.0011) $R^2 = 0.024$	6.222*** (0.0084) $R^2 = 0.503$	0.672*** (0.0173) $R^2 = 0.0252$

*** indicates significance at the 1 percent level.

Estimating equation (3), we first see from the first-step estimates in column six that Z is a very good instrument for the average absence in a neighborhood; the standard deviations are minuscule relative to the coefficients. In the second step, for the case where the behavior of private-sector employees is the dependent variable, we obtain γ = 0.581. According to this estimate, a typical private-sector individual has 0.581 more sick days if he lives in a neighborhood where the average number of sick days is one day higher than in another neighborhood. Similarly, a typical public-sector employee would have 0.762 more sick days. For the entire population (private plus public sectors), a person who lives in a neighborhood with an average that is one day higher than in another neighborhood would have 0.672 more sick days.

These estimates, standing alone, should be interpreted with caution. A possible reason for validity problems of the instrument is that private-sector workers with a preference for being absent are particularly likely to settle down in neighborhoods with a comparatively large share of public-sector employees, with perhaps greater social acceptance for work absence.

5. Movers within Sweden

So far, we have dealt with the reflection problem under the identifying assumption that private-sector individuals with preferences for absence do not self-select to neighborhoods with many public-sector employees – and a corresponding assumption for public-sector employees. In this section, we use two different identifying assumptions. In this section we consider individuals who have changed neighborhood within Sweden, and investigate whether they adjust their absence behavior to average behavior in the new neighborhood. Here, we can control for fixed individual heterogeneity since we have data on each individual’s behavior in the previous neighborhood. In the next section, we look at immigrants from abroad (in fact, mainly refugees). The self-selection problem is then mitigated since these individuals have to a considerable extent been allocated to neighborhoods by the authorities.

In the case of movers within Sweden, we limit the study to the individual's absence behavior during the first year in the new neighborhood (as compared to his previous behavior in the old neighborhood). Thus, we look only at very short-term adjustments.[14] Denoting the old neighborhood by *n* and the new by *m*, we estimate the following model:

$$S_{imt}^{mover} - S_{in,t-1}^{mover} = \alpha + \lambda_t + (X_{imt}^{'} - X_{in,t-1}^{'})\beta + \eta \cdot (\overline{S}_{m,t-1}^{non-mover} - \overline{S}_{n,t-1}^{non-mover}) + \varepsilon_{imnt}\,. \quad (5)$$

We use this analytical specification to investigate whether people who move from neighborhood *n* to neighborhood *m* adjust their behavior in response to the difference in average absence between these two neighborhoods. The coefficient η captures this influence. The reflection problem is avoided since there are two different population groups on the left-hand and right-hand sides of the equation; we can therefore estimate (5) by OLS. Moreover, with changes rather than levels on the left-hand side, thereby controlling for type, the selection problem is alleviated. The identifying assumption is that people who plan to *change* their absence behavior in the future do not tend to move to neighbourhoods with a particular level of average sickness absence.

The specification in (5) assumes that the adjustment is symmetric when moving between neighborhoods with different absence rates. In reality, individuals may be influenced differently when moving to neighborhoods with higher absence rates than when moving to neighborhoods with lower rates. We allow for this possibility in specification (5'):

$$\begin{aligned} S_{imt}^{mover} - S_{in,t-1}^{mover} &= \alpha + \lambda_t + (X_{imt}^{'} - X_{in,t-1}^{'})\beta + \eta \cdot (\overline{S}_{m,t-1}^{non-mover} - \overline{S}_{n,t-1}^{non-mover}) + \\ &+ \delta \cdot D_{it} \cdot (\overline{S}_{m,t-1}^{non-mover} - \overline{S}_{n,t-1}^{non-mover}) + \varepsilon_{imnt} \end{aligned} \quad (5')$$

where

[14] Long-term adjustment could in principle also be studied using panel data. However, some analytical complications would arise since individuals may move several times across neighborhoods.

$$D_{it} \equiv \begin{cases} 1 & \textit{if individual } i \textit{ has moved at time } t \textit{ to an area with} \\ & \textit{higher absensce, i.e., if } \bar{S}^{non-mover}_{m,t-1} > \bar{S}^{non-mover}_{n,t-1} \\ 0 & \textit{otherwise.} \end{cases}$$

The results are reported in Table 2. Since $\hat{\eta}$ of equation (5) is highly significant, we conclude that individuals to some extent adjust their behavior to average behavior in the new neighborhood, even in a very short time perspective. If someone moves to a neighborhood with one day's lower average absence, his absence falls by around 0.03 days. However, since $\hat{\delta}$ of equation (5') is not significant, there does not seem to be any asymmetry when moving to areas with lower sickness absence as compared to areas with higher absence.

Table 2: Movers within Sweden

	No. of individuals and observations	R^2	$\hat{\eta}$	$\hat{\delta}$
Symmetric specification (5)	1,551,059 ind. 2,202,466 obs.	0.0055	0.032*** (0.00678	—
Asymmetric specification (5')	1,551,059 ind. 2,202,466 obs.	0.0055	0.028*** (0.0094)	0.052 (0.0777)

Clearly, it does not seem likely that individuals make full adjustment to the average behavior in a new neighbourhood within a year. We therefore regard the results of Table 2 as an *indication* of social interaction, rather than a full quantification of such interaction. In this sense, the number 0.032 may be regarded as a lower bound on group influence.

6. Immigrants

We use the following model to investigate whether immigrants are affected by the work-absence behavior in the neighborhood where they settle down after arriving in Sweden:

$$S_{int}^{f} = \alpha + \lambda_t + X_{int}^{f\,'} + \gamma \bar{S}_{nt}^{s} + \varepsilon_{int} \,. \tag{6}$$

Here, S_{int}^{f} is the number of sick days of immigrant *i* in neighborhood *n*, while $\bar{S}_{nt}^{s}$ is the average number of sick days among native Swedes 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 either. We are thus able to rely on OLS, and we apply the identifying assumption 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").

Since we have data on each individual's country of origin, we can investigate whether immigrants with a cultural background similar to that of Swedes tend to adjust more than other immigrants to local Swedish absence behavior. The implied hypothesis is that such immigrants are likely to interact more than other immigrant groups with Swedes.

Since we want to study the transmission to immigrants of norms held by natives, it is natural to exclude neighborhoods where immigrants form a majority of the population. Indeed, we confine the regression to neighborhoods where the fraction of immigrants is less than 20 percent of the total population. The results are shown in Table 3. According to these highly significant estimates, sickness absence among immigrants is 0.629 days higher in a neighborhood where average absence among Swedes is one day higher than in another neighborhood. We interpret this figure as a proper estimate of the coefficient γ in equation (2). It is noteworthy that the order of magnitude of this estimate is about the same as in the IV estimate reported in Table 2.

The coefficients are particularly large for immigrants from other Nordic and EU countries (0.651 and 0.461 days, respectively). Our interpretation is that cultural affinity between immigrants and natives makes it easier to build networks in the new country. This result supports our hypothesis that social interaction helps explain individual sickness absence, since such interaction often takes place within networks. The effects are stronger, the tighter is the network.

Could the results reported in Table 3 depend on selection rather than on social interaction, thereby violating our basic identifying assumption? We could think of (at least) two types of such selection. One would be that the authorities (perhaps unintentionally) allocate immigrants with a high propensity to be absent from work to neighborhoods where the absence rates are particularly high among Swedes. This mechanism seems quite far-fetched, however. Self-selection by the immigrants themselves may be a more serious problem. We cannot fully rule out the possibility of some indirect mechanism by

Table 3: Estimates of γ in equation (6)

	All immigrants	
Region	Number of ind. and obs.	Estimate of γ
All regions	720,742 ind. 3,376,753 obs.	0.629*** (0.0063)
Nordic countries	210,059 ind. 1,088,923 obs.	0.651*** (0.0174)
EU (except Nordic countries)	77,982 ind. 358,797 obs.	0.461*** (0.0242)
Europe (except EU)	154,378 ind. 744,440 obs.	0.126*** (0.0176)
Africa	38,422 ind. 163,554 obs.	0.091*** (0.0308)
North America	21,655 ind. 92,321 obs.	0.278*** (0.0360)
Latin America	36,556 ind. 167,644 obs.	0.345*** (0.0347)
Asia	173,447 ind. 723,644 obs.	0.237*** (0.0157)
Oceania	3,626 ind. 14,146 obs.	0.222*** (0.0766)
Former Soviet Union	4,398 ind. 22,566 obs.	0.037 (0.0949)

which immigrants with a strong propensity to call in sick by self-selection would wind up in areas with many Swedes having the same propensity. For instance, immigrants with high labor-market ambitions may exhibit a particularly strong tendency to avoid areas with a weak labor market (to the extent that they are able to move at all). As a result, less ambitious immigrants might remain in areas where Swedes also have modest labor-market ambitions. If labor-market ambitions are negatively correlated with the propensity to call in sick, and if these ambitions are not reflected in the X vector, such a correlation may create a selection bias in the regression. Taking this possibility seriously, we have run regressions confined to recent immigrants (individuals who have lived in Sweden for one, two or three years, respectively). The results of these regressions are shown in Appendix 2; they imply that the longer an individual has been in Sweden, the higher is the coefficients γ. There are at least two possible interpretations of this finding. One is that it takes time for immigrants to observe and adjust to the behavior of native Swedes in the new location, and hence to build up and be influenced by networks with natives. Another interpretation is that the longer an immigrant has been in Sweden, the more likely is self-selection bias.

In summary, there are indications that social interaction between natives and immigrants matters for the sickness absence behavior of the latter. These indications are strengthened by the observation that the quantitative effects differ depending on the cultural background of the immigrants, and hence on the strength of their networks with natives.

7. Interaction between Neighborhood and Workplace Networks

If two individuals meet not only in their neighborhood, but also at their workplace, the strength of their social interaction may be accentuated; individuals would “rub shoulders” not only during their leisure time, but also during their working time. More generally, it is reasonable to assume that social norms and attitudes are transmitted more easily when individuals have more than one network in common. This form of interaction between two different networks can be used when estimating the effect of social interaction on work absence. For this purpose, we use the following model:

$$S_{iwnt} = \alpha + X_{inwt}^{'}\beta + \nu \cdot (CA_{inwt} \cdot \overline{S}_{nt}) + \lambda_t + \kappa_w + \mu_n + \varphi CA_{inwt} + \varepsilon_{inwt}\,, \qquad (7)$$

where the subscript *w* denotes the workplace. CA_{inwt} is a measure of the additional strength of the network facing individual *inw* at time *t* when he belongs to two different networks. It is defined as the fraction of the individual's neighbors who are also his coworkers. The parameters λ_t, κ_w and μ_n are fixed effects for year, workplace and neighborhood, respectively.[15]

Thus, the estimate of ν in equation (7) tells us whether there is an *additional* network effect for individuals who are not only neighbors but also coworkers. The coefficient ν, therefore, represents only a small fraction (an accentuation) of total social interaction. Note that the specification in (7) implies that the fixed neighborhood effect, μ_n, and the fixed workplace effect, κ_w, control for omitted variables in the *X* vector. In addition to the fixed effects and the interaction term $CA_{inwt} \cdot \overline{S}_{nt}$, equation (7) includes the density (concentration) measure CA_{inwt} separately. This allows us to control also for the possibility that the strength of the network in itself may be correlated with unobservable characteristics systematically related to the propensity to be absent from work. Our identifying assumption then is that there is no correlation between the interaction term $CA_{inwt} \cdot \overline{S}_{nt}$ and any remaining non-observable variables that affect sickness absence, i.e., $E\left(\varepsilon_{inwt} \mid CA_{inwr} \cdot \overline{S}_{nt}, \overline{S}_{nt}, X_{inwt}, \mu_n, \kappa_w \lambda_t\right) = E\left(\varepsilon_{inwt} \mid \overline{S}_{nt}, X_{inwt}, \mu_n, \kappa_w \lambda_t\right)$.

[15] Equation (7) has basically the same analytical structure as the one used by Bertrand et al. (2000) when studying the interaction between language groups and neighborhoods in an analysis of the reliance on social assistance ("welfare" in U.S. terminology) among ethnic minorities in the United States.

Note here that the vector X_i in (7) is a subset of the previously used X vector. The reason is that the neighborhood and workplace variables in X become redundant because the neighborhood and workplace fixed effects are included separately in the regression equation. The network-intensity variable only varies on the neighborhood/workplace level; we therefore adjust the standard errors for clustering within these cells (see e.g. Moulton, 1986).

When computing the ratio CA_{inwt}, we include in the denominator not only employees in the private and the public sector, but also employees in a third, ”unspecified” sector,[16] and self-employed persons. The reason is that CA_{inwt} is supposed to measure the probability of meeting a coworker in one's neighbourhood.[17]

Table 4 shows the results from the OLS estimation. The $\hat{\nu}$ is significantly different from zero. This means that we find evidence of social interaction on the utilization of the sick-pay insurance program. The size of the coefficient does not have a straightforward interpretation. But it is possible to calculate the marginal effect with respect to changes in average utilization of sick-pay insurance in the neighborhood, i.e., $\partial S_{iwn} / \partial \overline{S}_n$, which is easily seen to be equal to $\nu \cdot CA_{inwt}$. This parameter tells us how an increase in the average absence $\overline{S}_{nt}$ in a neighborhood influences individual absence through the interplay between neighborhood and workplace networks.

Table 4 shows that if the average absence in an average Swedish neighborhood increases by ten days, the strength-of-network effect adds 0.502 days to the average individual's absence. Following the discussion above, it should be stressed that this marginal effect is not comparable to our previous estimate of the γ coefficient in equation (2), which captures the full effect of a change in the average absence rate in the neighborhood.[18]

---

[16] This sector is rather heterogeneous and consists of persons with very weak ties to the labor market, i.e., they sometimes work temporarily in the private and/or public sector.
[17] The estimates do not change self-employed persons in the neighborhood are also included in the denominator.
[18] In the context of equation (7), social interaction is also reflected in the fixed effects μ_n and κ_w.

Table 4: The strength-of-network effect

Number of obs. and ind.	$\hat{v}$	$\partial S_{iwn} / \partial \overline{S}_n = \hat{v} \cdot \overline{CA}$	R^2
24,449,603 obs. 4,693,560 ind.	2.146*** (0.0273)	0.0502	0.012

Note: The numbers of observations and individuals in this table are somewhat smaller than the corresponding numbers in Table 1. The reason is that for each individual, we have deleted the individual himself from the data when computing the averages $\overline{S}_n$. For some neighborhoods, there is only one individual who works in each workplace; these cases therefore do not appear in the regression.

8. Concluding Remarks

We have used four different strategies for estimating the effects of social interaction within neighborhoods on absence behavior. All of these strategies unambiguously indicate that such interaction effects do in fact exist. However, all four strategies do not ask exactly the same question. In some cases we try to estimate the size of group effects on individual behavior (the parameter γ in equation 2), while in other cases we merely attempt to find *indications* of interaction effects. Moreover, we apply different identifying assumptions under the four strategies; this partly explains why the estimates differ. These approaches may be summarized as follows:

1. The private- vs. public-sector models in Section 4 rely on the identifying assumption that individuals with specific absence propensities do not self-select into neighborhoods on the basis of the share of public-sector versus private-sector employees in these neighborhoods.

2. The analysis of movers within Sweden in Section 5 relies on the assumption that individuals who expect to *change* their absence behavior do not choose to move to neighborhoods with particular average absence rates.

3. Our analysis of absence behavior among immigrants in Section 6 relies on the assumption that immigrants with particular propensities for absence do not settle down (as a result of administrative discretion or by self-selection) in neighborhoods where native Swedes have similar propensities.

4. The model exploiting interaction between neighborhood networks and workplace networks in Section 7 relies on the assumption that there is no correlation between neglected unobservable variables and the term for the network interaction.

It is noteworthy that the two attempts to quantify group effects (the IV model in Section 4 and the immigrant model in Section 5.2) yield rather similar results; the point estimates of the group effect, γ , are 0.672 and 0.629, respectively. If our identifying assumptions under these approaches (points 1 and 3 above) are not satisfied, the estimates would be biased upwards. It is also worth noting that the estimates of the group effect, γ, turn out to be higher in the case of immigrants from countries that are culturally close to Sweden (such as the other Nordic countries). This by itself could be interpreted as support for the presence of social interaction at the neighborhood level, irrespective of the validity of the identifying assumption.

The identifying assumptions under strategies 2 (movers within Sweden) and 4 (interaction between networks) may be less demanding. The results from these two approaches may therefore be more robust. Here, however, we do not obtain any estimate of γ, but only an *indication* of social interaction. The point estimate from the analysis of movers, according to which one additional day of average absence in the neighborhood leads to an increase of 0.032 days the year after a move, should be interpreted as a lower bound. After all, it is likely to take more than one year after a move for the full effect to show up. The estimated effect of the interaction between neighborhood and workplace networks, 0.0502, may be regarded as another estimate of the lower bound, since it does not capture group effects *within* these networks.

In this paper we have only dealt with a few types of group effects, i.e., personal contacts within neighborhoods (and the possible accentuation of such contacts through workplaces). There are, of course, other important channels for group effects. One example is personal interaction outside the neighborhood, for instance with relatives, those in the same profession, or those with similar interests. Another channel for group effects is through local and national mass media. Disregarding all these channels is, in itself, a source of underestimation of group effects on individual behavior.

References

Bertrand, Marianne, Erzo F. P. Luttmer and Sendhil Mullainathan, 2000, ”Network Effects and Welfare Cultures”, *Quarterly Journal of Economics*, 115 (3), 1019-55.

Bokenblom, Mattias and Kristin Ekblad, 2007, “Sickness Absence and Social Interactions – Evidence from a Swedish Municipality”, mimeo, Örebro University.

Duflo, Esther and Emmanuel Saez, 2003, “The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment,”, *Quarterly Journal of Economics* 118 (3), 815-842.

Glaeser, Edward L., Bruce I. Sacerdote and José A. Scheinkman, 2003, “The Social Multiplier”, *Journal of the European Economic Association* 1 (2), 345-353.

Greene, William H. (2003): *Econometric Analysis*, Fifth edition. Upper Saddle River, N.J.: Prentice Hall.

Henrekson, Magnus, and Mats Persson, 2004, “The Effects on Sick Leave of Changes in the Sickness Insurance System”, *Journal of Labor Economics*, 22 (1), 87-114.

Hesselius, Patrik, Per Johansson and Laura Larsson, 2005, ”Monitoring Sickness Insurance Claimants: Evidence from a Social Experiment”, IFAU, Uppsala University, Working Paper 2005:15.

Johansson, Per and Mårten Palme, 2005, “Moral Hazard and Sickness Insurance”, *Journal of Public Economics* 89 (9-10), 1879-1890.

Lindbeck, Assar, Sten Nyberg and Jörgen W. Weibull, 1999, ”Social norms and Economic In centices in the Welfare Stare”, *Quarterly Journal of Economics*, 114 (1), 1-35.

Lindbeck, Assar and Mats Persson, 2006, A Model of Income Insurance and Social Norms, CESifo Working Paper No. 1675.

Manski, Charles F., 2000, “Economic Analysis and Social Interaction”, *Journal of Economic Perspectives*, 14(3), 115-136.

Moffit, Robert, 1983, “An Economic Model of Welfare Stigma”, *American Economic Review*, 73 (5), 1023-1035.

Moulton, Brent, 1986, “Random Group Effects and the Precision of Regression Estimates”, *Journal of Econometrics* 32(3), 385-397.

Palmer, Edward (ed.), 2006, *Sjukförsäkring. Kulturer och attityder*. Försäkringskassan Analyserar 2006:16, Stockholm.

Statistics Sweden (2005): “Geography in Statistics – Regional Division in Sweden”. Mis 2005:2, Statistics Sweden, Stockholm.

Appendix 1: Explanatory variables in the *X* vector

For the individual	*Age* (all ages from 18 to 64, one dummy for each age, i.e., 46 dummies) *Education* (seven levels, one dummy for each level, i.e., six dummies) *Gender* (one dummy) *Marital status* (single, married/cohabitating, divorced; two dummies) *Has children aged 3 or younger* (one dummy) *Region of origin* (Sweden, Northern Europe, rest of Europe, etc.; 10 dummies)
For the workplace	*Industry* (60 industries, i. e., 59 dummies) *Sector* (central government, state-owned enterprise, local government, local government-owned enterprise, private firm, etc.; 11 sectors, i.e., 10 dummies)* *Size of workplace* (21 dummies: 1 employee, 2-10, 11-20, 21-30, …, 91-100, 101-200, 201-300, …, 901-1000, 1001-9999 employees)
For the neighborhood	*Urban or rural* (one dummy) *Life expectancy in the municipality* (average, gender-specific life expectancy among the 291 municipalities in Sweden) *Local unemployment* (expressed as the incidence of unemployment, i.e., the fraction of the labor force in the neighborhood that has received unemployment compensation at least once during the year. 19 dummy variables, one for each 5-percent interval)

* The distinction between industry and sector is that the former refers to the type of product or service produced, while the latter refers to ownership characteristics.

Appendix 2: Estimates for immigrants from abroad, according to length of stay in Sweden.

	All immigrants		Recent (three-year) immigrants		Recent (two-year) immigrants		Recent (one-year) immigrants	
Region	Number of observations	Estimate of γ	Number of observations	Estimate of γ	Number of observations	Estimate of γ	Number of observations	Estimate of γ
All regions	3,376,753	0.629*** (0.0063)	239,314	0.055*** (0.0070)	214,932	0.037*** (0.0053)	139,338	0.007** (0.0029)
Nordic countries	1,088,923	0.651*** (0.0174)	45,238	0.074*** (0.0245)	41,952	0.045 (0.0191)	28,995	0.002 (0.0107)
EU (except Nordic countries)	358,797	0.461*** (0.0242)	36,408	0.063*** (0.0198)	32,452	0.053*** (0.0180)	20,407	0.009 (0.0085)
Europe (except EU)	744,440	0.126*** (0.0176)	45,831	0.028 (0.0183)	41,013	0.009 (0.0143)	26,336	-0.005 (0.0091)
Africa	163,554	0.091*** (0.0308)	13,980	-0.023 (0.0324)	12,505	-0.010 (0.0223)	8,117	-0.006 (0.0178)
North America	92,321	0.278*** (0.0360)	12,269	0.060*** (0.0382)	11,123	0.035 (0.0272)	7,124	0.024* (0.0136)
Latin America	167,644	0.345*** (0.0347)	9,380	0.121** (0.0433)	8,376	0.130** (0.0525)	5,400	0.034 (0.0387)
Asia	723,644	0.237*** (0.0157)	72,525	0.020** (0.0099)	64,208	0.012 (0.0082)	40,865	0.005 (0.0030)
Oceania	14,146	0.222*** (0.0766)	3,017	0.125 (0.0895)	2,711	0.077 (0.0671)	1,719	-0.026 (0.0138)
Former Soviet Union	22,566	0.037 (0.0949)	399	-0.223 (0.4034)	361	0.189 (0.1535)	239	-0.022 (0.0333)
N/A	718	0.266 (0.3090)	267	0.163 (0.2113)	231	-0.005 (0.0045)	136	N/A

*** denotes significance at the 1 percent level, ** at the 5, and * at the 10 percent leve