The Strength of the Weakest Link: Sickness Absence, Internal Substitution and Worker-Firm Matching

Lena Hensvik and Olof Rosenqvist*

April 19, 2016

Abstract

We postulate that firms' production losses from absence depend on the employees' internal substitutability, incentivizing firms to keep absence low in positions with few substitutes. Using Swedish employer-employee data we show that absence is substantially lower in such positions even conditional on establishment and occupation fixed effects. The result reflects sorting on both entry and exit margins, with stronger separation responses when it was difficult to predict the absence of the employee beforehand. These findings highlight that internal substitution insures firms against production disruptions caused by absence and that absence costs are important aspects of firms' hiring and separation decisions.

*Hensvik: IFAU, Uppsala Center for Labor Studies (UCLS) and CESifo. Correspond via lena.hensvik@ifau.uu.se. Rosenqvist: Uppsala University and IFAU. We are grateful to Kevin Lang, Uta Schönberg, Karl-Oskar Lindgren, Illoong Kwon, Oskar Nordström Skans, Marcus Eliason and audiences at IFAU, UCLS, CReAM and EEA 2015 in Mannheim for helpful comments and suggestions.

1 Introduction

Many countries struggle with high sickness absence rates and large associated costs for firms.[1] Yet, little is known about how sickness absence affects key labor market outcomes, such as access to jobs, worker mobility and career trajectories. In addition, we know next to nothing about which strategies firms use to minimize the costs of employee absence.

This paper examines the idea that firms' production losses caused by temporary work absence depend on the internal substitutability of workers. Thus, firms should have incentives to keep absence low in jobs with few substitutes. Using Swedish administrative matched employer-employee data linked to information about individual sickness absence for almost 6 million worker-year observations, we document a robust positive relationship between employee absence and the number of internal substitutes defined by detailed occupations. The difference in the absence probability between more and less substitutable employees is substantial, even conditioning on establishment and occupation fixed effects: it is roughly equal to the average difference in work absence between young labor market entrants and middle aged workers or between workers with and without children.[2] This pattern holds irrespectively if we look at employees' own sickness absence or absences among parents caused by child sickness. Parents in jobs with few internal substitutes instead seem to shift part of their care leave for children to their spouse.

We then use several additional analyses to probe the mechanisms behind our results. About half of the effect remains after the inclusion of worker fixed effects, which indicates that sorting and on-the-job adjustments in absence

[1] It is reported that 131 million working days were lost due to sickness absence in the UK in 2013 (Office for National Statistics (UK), 2014). Another report from the UK estimates that employers pay GBP 9 billion (USD 12 billion) a year in sick pay and associated costs (Black and Frost, 2011). In Germany it is reported that employers spend about EUR 25 billion (USD 28 billion) per year on sick pay. This number is more than 1 percent of the total GDP in Germany (German Federal Statistical Office, 2011). Numbers for Sweden suggest that employers spent SEK 21 billion (USD 2.6 billion) on sick pay and associated costs in 2012 (Previa, 2013)

[2] The reported difference reflects the differential absence rate between employees with no internal substitutes and employees with more than five substitutes.

behavior are of equal importance for the observed association between sickness absence and internal substitutability. Further investigations of the selection mechanism show that workers hired to jobs with few internal substitutes have significantly lower pre-hire sickness absence than other new hires. They also display higher turnover rates caused by realized absence. Together these two results highlight that sorting occurs both via the entry and exit margin.

In addition, we find stronger selection effects weaker separation responses among hires with a strong pre-hire employment record; previous employment at another site the same firm or a coworker connection to an incumbent employee. We interpret this as suggestive evidence that sorting more pronounced when there was more information about the workers' absence-types ex ante, and that learning about match quality is an important determinant of turnover rates.

The paper contributes to several strands of the current literature. The idea that firms try to find the right employees for the right jobs is motivated by the notion that worker and firm heterogeneity can lead to match-specific gains in productivity.[3] But despite the theoretical foundations for match-specificity there is still little empirical evidence on cross-firm differences in hiring and the importance of worker-firm complementarities. One reason is that it is inherently difficult to measure, ex ante, how well a worker matches a particular job. Thus, researchers have mainly been restricted to infer the effects of match quality based on how wages and separations vary with tenure and job mobility (Nagypál, 2007; Lazear and Oyer, 2007).[4]

In addition, the discussion about match quality is often focused on complementarities in terms of worker skills (or human capital) and the skill requirements (or technology) of different jobs.[5] But it is equally likely to be impor-

[3]See Sattinger (1975), and Tinbergen (1956) for the original work on the problem of assigning heterogeneous workers to heterogeneous jobs.

[4]Two exceptions are Jackson (2013) who shows that teacher-school match effects explain a quarter of the variation in teacher quality and Fredriksson et al. (2015) who show that wages and job separations depend on how well workers cognitive abilities and personality traits match the abilities of the existing workforce.

[5]See, for example, Abowd et al. (2007) on how different components of skills are related to firms technological inputs; Andersson et al. (2009) on the relation between firms product

tant complementarities between other dimensions of employee attributes and firm technology that can affect the sorting of workers over jobs and, in turn, their subsequent labor market outcomes.[6] Our results highlight a previously overlooked, but seemingly important, dimension of match-quality related to complementarities between workers' absence rates and firms' possibilities of internal replacement.

A few recent papers specifically point to the importance of internal labor substitution for worker and firm outcomes. Jäger (2015) shows that internal labor markets are important for firms' replacements of sudden employee exits (caused by deaths), suggesting that firms face significant search-frictions in the external labor market. Our results complement his findings by highlighting the importance of internal substitution for insuring firms against temporary production disruptions caused by employee absence. In this sense, high-absence workers are weak links in jobs with few internal replacements.[7] Goldin and Katz (forthcoming) furthermore argue that the possibilities of employee substitution is a key factor behind the wage penalties associated with shorter hours, and in turn the gender pay gap.[8] The observed link between low internal substitutability and low probability of being absent to care for children is clearly consistent with their argument that the ease with which employees can substitute for each other affects individual absence costs.

market segment and the demand for worker innovation skills in the software industry or Lazear (2009) on firm-level heterogeneity in skill-weights.

[6]For example, Lazear (1998) argues that the match-quality between a worker and a given firm depends on the riskiness of workers, and firm-level characteristics such as expected time-horizon and the degree of private information.

[7]Our findings also relate to the studies documenting a positive association between sickness absence rates and firm size in the cross-section, which is consistent with the argument that production in small firms should be particularly sensitive to individual sickness absence (Barmby and Stephen (2000)), Dionne and Dostie (2007), Ose (2005) and Lindgren (2012). However, it is possible that this relationship also reflects other between-firm differences related to size. By exploiting variation in the number of substitutes within narrowly defined job cells, the present paper provides a more credible assessment of the direct relationship between sickness absence and employee substitutability.

[8]Their paper specifically looks at the pharmacist occupation and argues that enhanced substitutability (due to technological change and increased standardization) has decreased the wage penalty from shorter hours for women with children, and in turn the gender pay gap relative to other professions in the US labor market.

In addition, employee selection and hiring strategies is still somewhat of a black box (Oyer et al., 2011). Limited evidence suggests that employers are reluctant to hire applicants with a history of sickness absence, but remain uninformative of why (Eriksson et al., 2012). Our results suggest that firms' ability to internally substitute for absent workers is a key aspect of this process. But we also shed light on the role of information in the hiring decision. The fact that job separations respond to realizations of absence indicates that employment relationships are formed under uncertainty as in the seminal model of Jovanovic (1979). Consistent with several studies showing that firms rely on signals or informal search channels in order to screen for the right workers our findings suggest that pre-hire screening serves as a tool for firms to achieve an allocation of low-absence workers in unique positions.[9]

From the worker's perspective, these results imply that episodes of sickness absence affect the chances of accessing and retaining unique positions, which account for a non-trivial share of the labor market. Hence, workers have strong incentives to keep absence low in jobs with low internal substitutability, which they do by e.g. shifting the care for children to their partners.

The remainder of the paper is structured as follows. In section 2 we describe the data and clarify crucial definitions. The empirical specification and the results are presented in section 3. Section 4 concludes.

2 Data

2.1 Definitions and measurements

We use Swedish register data from 1997 to 2007. These data are drawn from registers administered by Statistics Sweden that follow all Swedish workers from 1985-2010, with unique person, firm and establishment identifiers. In the main analysis we restrict the sample to jobs in the private sector. The

[9]Empirical studies in this literature suggest that employers use observable signals such as education (Farber and Gibbons, 1996; Altonji, 2005; Lange, 2007; Schönberg, 2007), unemployment status (Eriksson and Rooth, 2014), and referral ability (Hensvik and Nordström Skans, forthcoming) to form expectations about prospective workers productivity.

reason is that the definition of the establishment is more precise in the private sector.[10] To these data we add demographics from a population-wide dataset and information on occupation codes, which is available from 1997-2010 for a large sample of private establishments covering almost 50 percent of private sector workers.[11]

2.1.1 Measuring internal substitutability

We define employee substitutability as the number of other workers within the same combination of establishment and occupation (ISCO-88, 3-digit level) in a given year. For example, an administrator at an establishment that in total employs four administrators will have three substitutes. In order to focus on regular workers, we drop employees in managerial positions. We also drop employees at very small establishments (less than three employees).

Our definition of employee substitutability is supported by Jäger (2015) who shows that when an employee exits (due to death) firms increase their demand for the remaining workers in the same, but not in other occupations as the deceased. This clearly indicates that firms regard employees within the same narrowly defined occupations as closer substitutes than employees in other occupations. In most specifications, we let an indicator for having 0-5 substitutes define low internal substitutability, but we also show results from more flexible models.

It is likely that the number of substitutes will be measured with error. Specialization within occupations could lead to an overstatement of the true level of substitutability. But it is also possible that some coworkers have overlapping skills even if they occupy different jobs, in which case we may understate the true number of substitutes. We address this issue in the robustness section

[10] In the public sector, all individuals that are employed by the same municipality are sometimes registered as belonging to the same establishment.

[11] We start the observation period in 1997 since this is the first year that we can observe occupations in our data. The reason for ending already in 2007 is that we in some cases want to follow workers for a 3-year follow-up period. In terms of sampling, a new random sample is drawn each year and the establishments are stratified by firm size and industry. Table A3 shows the distribution of establishments (col. 2) and employees (col. 3) with respect to establishment size.

using alternative definitions of employee substitutability that e.g. takes firm size into account.

2.1.2 Measuring sickness absence

We add sickness absence spells from the Swedish Social Insurance Agency. These data include all spells longer than two weeks.[12] Sickness absence will generally be defined as an indicator for having at least one such spell in a given year. But in some specifications we will also consider absence on the intensive margin, by using the log of sickness benefits as the outcome of interest.

The fact that we cannot observe shorter spells is obviously a limitation of the data, and we will therefore complement our analysis with short-term work absence due to care leave for sick children as an alternative outcome measure. In Sweden, parents with small children (0-10 years old) can be absent from work to care for sick children (that are too sick or infectious to be in school or in daycare).[13] The parent that stays home receives Temporary Parental Benefits from the Social Insurance system from day one meaning that these benefits data also pick up short term absence spells.[14]

2.1.3 Defining hires, pre-hire and realized absence

We examine the role of worker sorting in more detail using a dataset consisting of new hires. We define new hires as employees observed in an establishment in a given year, but not in the same establishment *or* in the same firm in any of the five preceding years. For each hire, we measure their pre-hire sickness absence as the average incidence of having at least one sickness absence spell

[12] The data include all spells for which the individual was entitled to sickness benefits from the social insurance system. Since spells shorter than two weeks are paid by the employers, these are not available in our data.

[13] 90 percent of all parents in Sweden have their children between 3-6 years of age in subsidized child care (Mörk et al., 2013)

[14] Parents may claim benefit compensation for up to 120 days per year. The replacement rate is 80 percent of lost earnings up to a monthly wage ceiling of SEK 37,000. The benefit compensation data contain information on the total amount of child sick benefits received each year, from which we construct an indicator for having at least one child sick spell in a given year.

(longer than two weeks) per year in the three years prior to employment. In order for all new hires to have at least three pre-hire years, we restrict the sample to workers with at least 4 years of labor market experience.[15] We will also examine the probability of job separation when the worker-absence-type is revealed. To this end we define realized absence of new hires as the average sickness absence probability in the hiring year and the year after entry.

2.1.4 Measuring uncertainty

Part of our empirical analysis aims to contrast realized matches between workers and firms where the hiring decision was based on more or less information about worker-absence-type. To this end, we use three different proxies for the amount of information about the employees in the matching stage:

i. *Pre-hire employment*: an indicator for if the new employee was employed in t-3, t-2 and t-1.

ii. *Firm connection*: an indicator for if the new employee was employed in the same firm, but in another establishment, sometime between t-1 and t-5.

iii. *Coworker connection*: an indicator for if the new employee was ever employed in the same establishment (at another firm) as at least one of the incumbent employees of the hiring establishment.[16]

These information proxies are all based on the notion that the employment history of a worker provides information about his future absence behavior. Hence, when new hires fulfill one of the three criteria above, we assume that the hiring decision was based on a more precise signal about the prospective hire's absence-type.

[15] In other words, labor market entrants with less than 4 years since they graduated from their highest education are excluded.

[16] We construct dyads for each hire-incumbent combination (i.e. if a new worker comes to an establishment with 10 incumbent workers we create 10 dyads). For each dyad we add information on the full history of employers for both agents back to 1985. A coworker connection is defined as having overlapping employment spells in the same establishment.

Although it is clear that these measures are far from perfect, several studies support our choice of information proxies. Work by Eriksson and Rooth (2014) shows that employers are reluctant to hire people from non-employment, which indicates that non-employment is associated with some degree of uncertainty about worker type. Thus, we find it reasonable to expect that there is more information available about workers with a strong attachment to the labor market.[17] Schönberg (2007) further shows that hard-to-observe characteristics of college graduates are more easily assessed by the current firms than by outside firms. Under this assumption, we expect that matches involving workers with an earlier connection to the recruiting firm are based on better information about the worker-absence-type.

Finally, there is recent evidence that incumbent employees can provide valuable information about the productivity of prospective hires with whom they have worked in the past (Dustmann et al., 2015; Hensvik and Nordström Skans, forthcoming). Based on these findings, we assume that firms' can make better predictions about the absence-type of former coworkers to their current employees.

2.2 Descriptive statistics

Tables A1 and A2 in the Appendix contain descriptive statistics on the sample of all workers and on the sample of new hires respectively. There are 6 million observations in the full sample (Table A1) and 400,000 new hires (Table A2). About 20 percent of these occupy jobs with 0-5 substitutes, which suggests that positions with low employee substitutability account for a significant share of the labor market.[18] About 4 percent of the workers have truly unique jobs (i.e. 0 substitutes) and about 11 percent of the employees have at least one sickness

[17] Farber and Gibbons (1996), Altonji (2005) and Lange (2007) show that employers overprice formal credentials (and underprice hidden talents) among inexperienced workers, which further supports that there is less information about worker type for employees with weaker labor market experience.

[18] 1,050,017 (73,366) out of the 5,863,497 (387,901) workers (hires) have jobs with 0-5 substitutes

absence spell that is longer than two weeks in a given year.[19] Consistent with our hypothesis, the incidence of sickness absence is lower for workers with relatively few substitutes (0-5), but these workers differ in other aspects as well; they are for example employed in smaller establishments and in more skilled professions with higher wages suggesting that they have key positions within the firms.[20] Workers in relatively unique positions are in addition older and more often women, although education levels appear similar as to other employees.

The image of the new hires is very much in line with the full sample. Importantly, positions without substitutes are present in all occupational skill levels (the note to Table A2 gives the distribution). In contrast to our hypothesis however, the pre-hire sickness absence rate is higher for workers who enter relatively unique positions, while wages are about the same. But as noted before, it is important to account for other aspects that differ systematically between individuals in more/less unique positions before we can draw conclusions about the relationship between employee absence and internal substitutability.

3 Empirical strategy and findings

3.1 Empirical specification

We start by exploring the association between present sickness absence and the number of employee substitutes among all private sector workers by estimating Eq. (1) by OLS:

$$A_{ijpt} = \gamma S_{ijpt} + \alpha_j + \alpha_p + \theta_t + \beta X_{it} + \delta Z_{jt} + \epsilon_{ijpt} \tag{1}$$

where the outcome A_{ijpt} is the incidence of sickness absence for worker i in establishment j and profession p in year t. S_{ijpt} measures employee substi-

[19] The figure on sickness absence is confirmed by estimates from Statistics Sweden (Statistics Sweden, 2007).

[20] The summary statistics show the distribution of workers/hires across a broader set of occupations (1-digit level). When defining the number of substitutes we use more detailed occupation codes (3-digit level).

tutability within each job, defined by the interaction between the establishment and the 3-digit occupation code.[21] α_j and α_p are establishment and occupation fixed effects respectively. We also include year fixed effects, θ_t to account for e.g. business cycle swings potentially correlated with firms' organization of work and individual sickness absence. The worker characteristics X_{it} consist of gender, age, education, country of origin and an indicator for having children under the age of three.[22] Finally we include establishment size Z_{jt}. ϵ_{ijp} is the error term.

The parameter of interest is γ, which aims to capture the relationship between the number of internal substitutes and work absence.[23] It should be noted that the model is fairly rich as it accounts for unobserved characteristics of both occupations and establishments that could generate a spurious correlation between employee substitutability and absence.

We also want to disentangle to which extent γ captures behavioral responses and/or employee selection on the entry and exit margin. As a first step we therefore add worker fixed effects to Eq. (1), which means that we account for the selection of workers over jobs with few/many substitutes. Second, we estimate Eq. (1) separately for new hires and replace the outcome with an indicator for the *pre-hire* sickness absence, defined as the average incidence of having at least one sick leave spell longer than two weeks per year in the three years prior to entry. Since pre-hire sickness absence is potentially correlated with past employment we also control for the employment probability in the same time period. Finally, we examine the separation response to realized sickness absence among new hires by estimating the following equation:

$$Separation_{ijpt+2} = \delta \overline{A}_{ijpt} + \alpha_j + \alpha_p + \theta_t + \beta X_{it} + \delta Z_{jt} + \epsilon_{ijpt} \qquad (2)$$

[21] In our main specifications, we consider jobs with 0-5 substitutes in the same occupation and establishment as jobs with low substitutability. Sometimes we also use models with slightly different specifications, which we then state clearly.

[22] We group individuals by their country of origin into the following six categories: Sweden, rest of the Nordic countries, rest of Europe, North America, South America, and the rest of the world

[23] The baseline analysis focuses on sickness absence on the extensive margin. As a robustness check we also consider outcomes that capture the intensive margin.

where $Separation_{ijpt+2}$ is an indicator for if worker i hired in year t left establishment j between year $t+1$ and $t+2$, and $\overline{A}_{ijpt}$ is the realized absence of entrant i measured as the averaged incidence of absence over the entry year (year t) and the first year into the employment spell ($t+1$) (we focus on entrants that stayed for at least one year in order to be able to observe their realized sickness absence). The controls are the same as in Eq. 1 and ϵ_{ijp} is the error term.

The aim of δ is to capture the separation response to the realized absence behavior among newly hired workers. To examine whether this response depends on the internal substitutability of employees we also estimate versions where the model in Eq. 2 is fully interacted with employee substitutability (i.e. with S_{ijpt}, in Eq. 1).

3.2 Baseline results: employee substitutes and absence

Figure 1 shows the estimates from Eq. (1) when we include dummy variables for having up to 5 substitutes (employees with more than 5 substitutes constitute the reference category). The estimates are all statistically negative on the 1-percent level, ranging between 1 and 2 percentage points. Hence, workers with few close substitutes have lower absence rates. Interestingly, the estimates become smaller in absolute value as the number of employees performing the same job increases, which is consistent with the idea that the costs of employee absence, in terms of production disruptions, increase as the possibilities of internal substitution decrease.

In the Appendix (Figure A1) we show the same relationship for up to 10 substitutes (employees with more than 10 substitutes constitute the reference category). These results suggest that there is a significant jump in the absence probability when the number of substitutes increases from 0 to 1. Beyond that, there is a fairly linear relationship between employee absence and the number of substitutes.[24] The magnitudes of the estimates are substantial,

[24]The difference in absence probability between jobs with ten and more than ten substitutes is around 0.5 percentage points. This remaining difference may seem surprising but is probably due to the fact that we have measurement error in the possibilities of replacing an

especially for the coefficients on 0 and 1 substitutes: the difference in sickness absence between jobs with more than 5 substitutes and jobs with 0 substitutes, conditional on the model, is roughly equivalent to the estimated difference in absence rates between workers in their 20s and 40s, or between workers with and without small children (0-3 years of age).

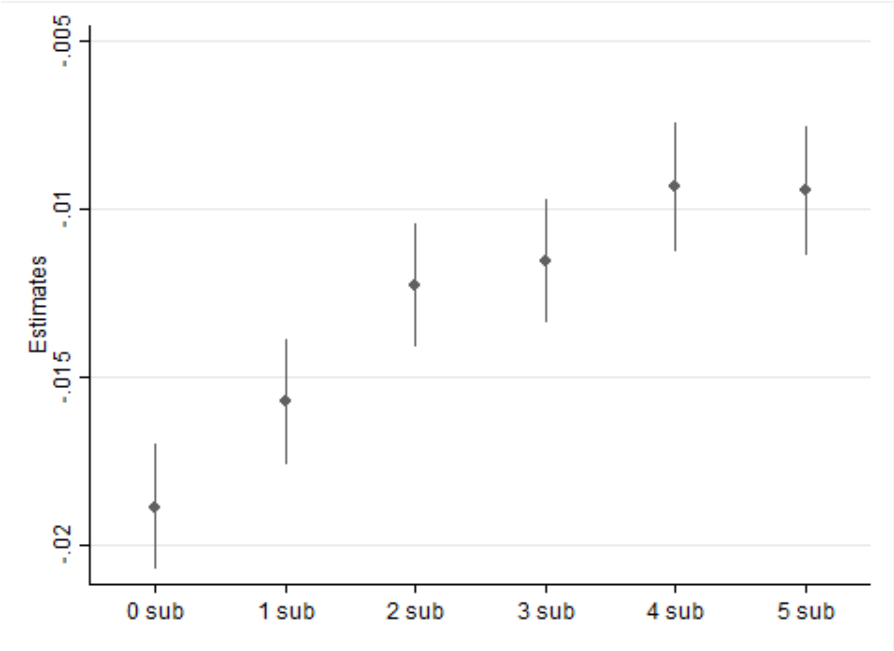

Figure 1: Probability of absence and the number of employee substitutes
Notes: The figure shows the the estimated coefficients on dummies for 0-5 substitues in Eq. 1. The reference category is employees with more than 5 substitutes and the background controls are gender, age, education, birth country, having children aged 0-3 and establishment size. The model also includes year, occupational, and establishment fixed effects. Standard errors are clustered on the establishment level.

Table 1 shows the point estimates (with and without worker characteristics) when we for simplicity only use an indicator for low substitutability, defined as having 0-5 substitutes. As before the reference category is employees with more than 5 substitutes. Overall, these results show a strong negative correlation between low internal substitutability and work absence.

absent employee which is likely to decrease with the size of the job-cell (in large cells there is a greater chance that at least some workers are perfect substitutes for each other).

Table 1: Sickness absence and internal substitutability

Outcome: Sickness absence in t	(1)	(2)
Low substitutability	-0.0104***	-0.0131***
	(0.0006)	(0.0006)
Number of observations	5,863,497	5,863,497
Mean of dependent variable	0.109	0.109
Background controls	No	Yes
Year fixed effects	Yes	Yes
Occupation fixed effects	Yes	Yes
Establishment fixed effects	Yes	Yes

Notes. The standard errors are clustered on the establishment level. Low substitutability is defined as having 0-5 substitutes (i.e. the reference category is employees with more than 5 substitutes). The background controls are gender, age, education, birth country, having children aged 0-3 and establishment size. $*$, $**$, and $***$ denote statistical significance at the 10-,5-, and 1-percent level.

3.3 Evidence from child sick spells

A limitation of our data is that we only observe absence spells longer than two weeks. To test if our results also extend to shorter absence spells we therefore include an alternative absence measure: *Care leave for sick children*, which also includes short-term work absence.[25] 65 percent of the parents have at least one absence spell according to our definition, which suggests that this type of absence is a first order concern for firms that employ workers with small children in the household.[26]

We restrict this analysis to individuals with at least one child between 0 and 10 years old (these are the children that parents are entitled to be at home with) and use an indicator for having positive Temporary Parental Benefits in a given year as the outcome. To see if employee substitutability affects the division of care for sick children within the family we also use the corresponding absence measure for the partner as an outcome (the sample is

[25]The reason is that parents receive benefits from the Social Insurance System from day one.

[26]Following women in Sweden who had their first child in 1994, Boye (2015) shows that the average women is absent from work for 5 days per year and the average man is absent 2.5 days per year during the first 10 years of the child's life, with higher absence rates for children in daycare ages.

then further restricted to individuals with a cohabiting partner).

Table 2 presents the results from this exercise using the model described by Eq. (1). The estimate in column (1) clearly suggests that workers in jobs with few substitutes are significantly less likely to be absent due to care leave for sick children. Thus, the results are in line with our general findings in Table 1, although compared to the baseline they are smaller in magnitude.[27] Interestingly, the partners of employees with few substitutes are more likely to be home caring for sick children (see column [2]) and the magnitude of the estimate is almost equal to the estimate in column (1).[28] Thus, children to workers with few employee substitutes are no less sick than other children, instead these workers seem to avoid work interruptions by shifting work absence to their partners.[29]

3.4 Robustness checks

3.4.1 Alternative measures of employee substitutability

Our baseline measure of internal substitutability is the number of employee substitutes in the same occupation. But it is well possible that the substitutability of workers could interact with the size and the organization of the establishment. We may for example overstate the degree of substitutability in large establishments if employees are organized in different departments that make substitution difficult. More coworkers in general could at the same time imply that employees are more substitutable as there is a higher likelihood that some workers have overlapping skill sets even though they occupy

[27] The weaker relationship may reflect that a large share of parents have at least one child sickness spell, which is likely to make the extensive margin less relevant. In Table A7, which we discuss below we show that the effect on the intensive margin is very similar as for own sickness absence

[28] The estimate in column (1) is almost identical if we use the same sample as in column (2) (i.e. employees with cohabitating partners).

[29] As a robustness exercise, we have also looked at the relationship between own substitutability and the partner's *own* sickness absence. The estimate is close to zero and precisely estimated (-0.0020 [0.0010]). The fact that the partner's response is concentrated to child leave days is reassuring, as these (but not own sick leave days) can be shifted between parents.

Table 2: Evidence from child sick spells

	(1)	(2)
Outcome:	Absence to care for sick child	
	Own absence	Partners' absence
Low substitutability	-0.0115***	0.0087***
	(0.0017)	(0.0018)
Number of observations	1,911,734	1,767,118
Mean of dependent variable	0.654	0.553
Background controls	Yes	Yes
Year fixed effects	Yes	Yes
Occupation fixed effects	Yes	Yes
Establishment fixed effects	Yes	Yes

Notes. The standard errors are clustered on the establishment level. Low substitutability is defined as having 0-5 substitutes (i.e. the reference category is employees with more than 5 substitutes). The background controls are gender, age, education, birth country, having having children aged 0-3 and establishment size. In column (1) we restrict the sample to individuals with at least one child younger than 11 years of age and in column (2) we further restrict the sample to individuals with cohabiting partners. We further control for the number of children in the following categories: 0-3 years, 4-6 years and 7-10 years. $*$, $**$, and $***$ denote statistical significance at the 10-, 5-, and 1-percent level.

different jobs.

It is therefore not clear how to (and if we should) adjust the number of substitutes to establishment size. As a starting point, column (4) of Table A3 shows how much of the identifying variation in the variable *Low substitutability* that comes from establishments of different sizes. The figures are based on the squared residuals from a regression of S_{ijpt} on the full covariate set in Eq. (1). It is clear that small to medium establishments account for a large share of the variation: 41 percent comes from establishments with 3-49 employees and 40 percent comes from establishments with 50-249 employees. To test how relevant our results are for establishments of different sizes we therefore reestimate the baseline model separately for those with 3-49, 50-249, 250-500 and more than 500 employees.[30] The estimates from this exercise are plotted in Figure A2. All four estimates are significantly negative on the 1 percent

[30]The division is based on a classification that Statistics Sweden uses when they collect data from firms.

level and the magnitudes of the estimates are roughly similar to the estimate presented in column (2) of Table 1. Thus, our measure of substitutability is relevant for both small and large establishments.

For completeness we have, however, also tested two other definitions of low substitutability based on the logic that employees are less substitutable in larger establishments (for a given number of substitutes). First, we define low substitutability when one of the following criteria is fulfilled: (i) no substitutes in establishments with 3-49 employees; (ii) <4 substitutes in establishments with 50-249 employees (iii) <7 substitutes in establishments with 250-500 employees or (iv) <10 substitutes in establishments with more than 500 employees. The second alternative definition is based on the number of substitutes divided by establishment size and defines low substitutability when this share is below the 0.03 (which corresponds to the tenth percentile).

These definitions are of course arbitrary but offers a way of relating the notion of substitutability to the overall size of the establishment (Columns [5-6] of Table A3 shows that more of the identifying variation now comes from larger establishments). However, when we reestimate Eq. (1) using these two alternatives we obtain virtually identical estimates (see Table A4).[31] Overall, we conclude that the link between employee substitutability and absence is relevant for establishments of all sizes (rather than only relatively small ones) and that our results are robust to different variations in the definition of internal substitutability.

3.4.2 Alternative explanations and specification checks

The strong association between sickness absence and substitutability naturally raises the relevant question whether wages differ between more and less substitutable jobs. Table A6 in the Appendix suggests that is indeed the case. We obtain these estimates by replacing sickness absence as the outcome in

[31] Column (4) also shows the linear relationship between the share of substitutes and absence, which suggests that a standard deviation increase in the share of substitutes is associated with 0.8 percentage points higher probability of absence (0.3*2.7).

Eq. (1) with the log of the monthly full-time wage.[32] The results in column (1) suggest that employees with 0-5 substitutes have 1.3 percent higher wage on average relative to employees with more than 5 substitutes, conditional on establishment and occupation fixed effects.[33]

This wage premium could both reflect that unique jobs are more productive, and/or that the employees in unique jobs have more productive (unobserved) skills that are correlated with their absence-type. In column (2) of Table A6 we show the baseline estimate when we, as a robustness check, hold the wage constant. Even if it is potentially problematic to control for the wage (as the wage is likely to be endogenous to the level of absence), it is reassuring to see that this only has a minor impact on the main estimates.

Column (3) of Table A6 shows the estimate when we add the public sector employees to our sample. This estimate is somewhat smaller, but still significant and of important magnitude suggesting that the relationship between the number of employee substitutes and sickness absence holds in the full economy. In column (4) of Table A6 we use data on private sector employees for the years 2005-2007. In these years the occupational code is available on a 4-digit level and thus we can test if our main results in Table 1, which are based on a 3-digit occupational code, are robust to finer definitions of occupations. The estimate is very similar to the one in column (2) of Table 1 and confirms the general picture of low sickness absence in jobs with few substitutes.

Finally, in Table A7 we use the log of annual sickness benefits instead of an indicator for sickness absence as the outcome, which picks up the length and number of absence spells. Conditional on being absent for at least two weeks, employees with fewer substitutes have roughly two percent fewer absence days

[32] The wage is the wage the employee had during the sampling week expressed in full-time monthly equivalents. The variable includes all fixed wage components, including piece-rate and performance pay as well as fringe benefits. Overtime pay or paid leave is however not included. The monthly wage is adjusted to full-time for part-time workers by Statistics Sweden. For the blue-collar workers the wage is typically obtained by the hourly pay rate times the number of hours that correspond to full-time employment. For the white-collar workers it reflects the September wage adjusted by the share of part-time work during the same month.

[33] Estimating the same model for new hires we find an identical wage premium.

(Panel A) and almost four percent fewer care leave days due to child sickness (Panel B), compared to employees with more substitutes (the received benefits are closely related to the number of leave days). The results thus seem qualitatively robust to variations in the way we measure sickness absence and suggest that workers with few substitutes have lower absence rates on both the extensive and the intensive margin.

3.5 Behavioral vs. entry and separation responses

Behavior and entry

The documented relationship between sickness absence and internal substitutability may both reflect a selection effect (systematic sorting into and out of jobs with few substitutes) and a behavioral effect (workers adjusting their absence behavior when they have few substitutes). To examine the relevance of these two explanations we exploit variation in the number of substitutes for the same worker over time by adding worker fixed effects to the baseline specification. The estimate presented in column (2) of Table 3 is roughly halved compared to the baseline estimate in column (1) but remains significantly negative on the 1-percent level. Hence, workers do adjust their work absence depending on the number of employee substitutes.[34] Taken together, these results suggest that the correlation between internal substitutability and sickness absence entails both a selection component and a behavioural component that appear to be of similar importance.

In the third column of Table 3 we replace present sickness absence with the *pre-hire* sickness absence described in Section 2.1 in a sample of new hires. Consistent with our earlier results, this estimate clearly suggests that workers hired into positions with fewer substitutes are more likely to be low-absence types. Reassuringly, this estimate (0.4 percentage points) is very similar to the difference between the estimates with and without worker fixed effects in the full sample, which supports the interpretation that workers with few absence

[34]This could reflect that employers spend more money on employee wellness for unique employees or increase the pressure not to be absent.

spells sort into jobs with low substitutability.

Table 3: Behavior vs. sorting into jobs

	(1)	(2)	(3)
Sample:	All workers	All workers	New hires
Outcome:	Present absence	Present absence	Pre-hire absence
Mechanism:	*Baseline*	*Behavior*	*Selection*
Low substitutability	-0.0131***	-0.0058***	-0.0043***
	(0.0006)	(0.0007)	(0.0014)
Number of observations	5,863,497	5,863,497	387,901
Mean of dependent variable	0.109	0.109	0.116
Background controls	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes
Establishment fixed effects	Yes	Yes	Yes
Worker fixed effects	No	Yes	No

Notes. The standard errors are clustered on the establishment level in columns (1) and (3) and on the worker level in column (2). Low substitutability is defined as having 0-5 substitutes (i.e. the reference category is employees with more than 5 substitutes). The background controls are gender, age, education, birth country, having children aged 0-3 and establishment size. In column (3) we also control for the pre-hire employment status of the new hire. $*$, $**$, and $***$ denote statistical significance at the 10-, 5-, and 1-percent level.

Separations

We have so far focused on employee absenteeism and the selection into jobs. But in Table 4 we complement the analysis by asking how the realized sickness absence among new hires (measured as the average sickness absence probability in t and t+1) affects (i) the probability of exiting the employment relationship as well as (ii) the probability of receiving more employee substitutes within three years after entry. We study the first question by estimating Eq. (2) using a sample of new hires that stay in the establishment for at least one year (in order to be able to observe realized sickness absence). The outcome is an indicator for exiting the establishment between t+1 and t+2 (Panel A).[35] We study the second question using a sample of new hires that are observed in

[35]As a robustness check we have also used an indicator for not being observed in the establishment in either t+2 or t+3. This does not substantially change the results.

the establishment at least until t+3. The outcome is an indicator for having more substitutes in t+3 than in t (Panel B).

The results suggest that higher realized sickness absence generally is associated with significantly higher turnover rates (Panel A, column [1]), and a higher likelihood of receiving more employee substitutes (Panel B, column [1]).[36] This relationship is particularly strong for workers employed in jobs with low internal substitutability (columns [2-4]), suggesting that sorting on the bases of sickness absence also occurs via the exit margin.

3.6 The role of information

The fact that job separations respond to realizations of sickness absence suggests that matches are formed under some remaining uncertainty. In this section we examine the direct role of information for the selection into and out of jobs. We use the information proxies described in Section 2 to assess the degree of uncertainty in the hiring stage: (i) an indicator for being employed in t-1 to t-3 (*Pre-hire employment*), (ii) an indicator for previous employment in another establishment within the same firm (*Firm connection*) and (iii) an indicator previous employment in the same establishment as an incumbent employee (*Coworker connection*).[37]

If employers are reluctant to hire applicants with an observable history of sickness absence we expect that better information should be associated with lower pre-hire absence among new hires. This is also what we see in the first row of panel A of Table 5: hires have between one and two percentage points lower pre-hire absence for two out of the three information measures. Importantly, the negative relationship between information availability and pre-hire sickness appears about twice as strong in jobs with few substitutes as in jobs

[36] Interestingly, when we condition on being observed in t+1 and t+2 and use the average sickness absence probability in t+1 and t+2 as an explaining variable for leaving the establishment in t+3 the estimate in Panel A, column (1), is substantially lower. This is consistent with the notion that the marginal effect of exhibiting bad properties (in this case high sickness absence), in relation to the job, on job separation probability should decrease with tenure (see Kwon (2005) for an interesting contribution on this topic).

[37] When we use the previous firm connection as the information proxy we relax the new hire definition and include new hires on the workplace with a history within the firm.

Table 4: Realized sickness absence and post-hire outcomes

	(1)	(2)	(3)	(4)
	All	Unique jobs	Not unique jobs	Diff.
Outcome:	*A: Separation in t+2*			
Realized absence	0.1100***	0.1284***	0.1072***	0.1072***
	(0.0037)	(0.0097)	(0.0040)	(0.0040)
Realized absence ×Low subst.				0.0212**
				(0.0105)
Observations	336,026	63,624	272,402	336,026
Mean of dependent variable	0.270	0.280	0.267	0.270
Outcome:	*B: More substitutes in t+3*			
Realized absence	0.0145*	0.0576**	0.0120	0.0120
	(0.0075)	(0.0249)	(0.0079)	(0.0079)
Realized absence ×Low subst.				0.0456*
				(0.0261)
Observations	110,869	18,838	92,031	110,869
Mean of dependent variable	0.487	0.446	0.496	0.487
Background controls	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes	Yes
Establishment fixed effects	Yes	Yes	Yes	Yes

Notes. In Panel A the sample is restricted to new hires that are observed in the establishment in t+1. Low substitutability is defined as having 0-5 substitutes (i.e. the reference category is employees with more than 5 substitutes). In Panel B the sample is restricted to new hires that are observed in the establishment in t+1, t+2 and t+3. The standard errors are clustered on the establishment level. The background controls are gender, age, education, birth country, having children aged 0-3 and establishment size. In column (4) all variables are interacted with the variable indicating 0-5 substitutes. $*$, $**$, and $***$ denote statistical significance at the 10-, 5-, and 1-percent level.

with many substitutes. Thus, when employers are recruiting to positions with few internal substitutes they react even stronger to information about worker-absence-type, which suggests that screening for low absence workers seems to be more important when there is low internal substitutability of workers.

Panel B shows how job separations induced by realized sickness absence are related to the information available in the hiring stage. Intuitively, separations should respond more to realized sickness absence behavior if there was

Table 5: The role of information

	(1)	(2)	(3)
Information proxy:	Pre-hire employment	Firm connection	Coworker connection
Outcome:	*A: Pre-hire sickness absence*		
Sample:	All new hires		
Better informed (baseline)	-0.0208***	-0.0093***	-0.0008
	(0.0015)	(0.0011)	(0.0010)
Better informed ×Low subst.	-0.0249***	-0.0065***	-0.0045
	(0.0043)	(0.0024)	(0.0029)
Observations	387,901	586,994	387,901
Mean of dep. variable	0.116	0.115	0.116
Outcome:	*B: Separation in t+2*		
Sample:	New hires with few subst.		
Realized absence (baseline)	0.1575***	0.1284***	0.1318***
	(0.0390)	(0.0097)	(0.0113)
Realized absence × Better informed	-0.0441	-0.0777***	-0.0481*
	(0.0404)	(0.0157)	(0.0273)
Number of observations	63,624	95,236	63,624
Mean of dependent variable	0.280	0.276	0.280
Background controls	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes
Establishment fixed effects	Yes	Yes	Yes

Notes. The standard errors are clustered on the establishment level. The background controls are gender, age, education, birth country, having children aged 0-3, pre-hire employment status (not included in column [1]) of Panel A or in Panel B) and establishment size. Panel A: All variables are interacted with the indicator for low substitutability, defined as having 0-5 substitutes (i.e. the reference category is employees with more than 5 substitutes). Panel B: All variables are interacted with the information proxy. The sample corresponds to the sample in column (2) of Panel A in Table 4. In column (2), we relax the new hire definition and include new hires with previous employment in another establishment within the same firm, which explains why the sample size is larger than in columns (1) and (3). $*$, $**$, and $***$ denote statistical significance at the 10-, 5-, and 1-percent level.

less information about absence type beforehand. For simplicity, we restrict this analysis to jobs with few substitutes (≤ 5) and interact realized absence with our information proxies. Consistent with the results in Panel A of Table 4, there is a strong relationship between realized sickness absence and the probability of job separation. However, this relationship is weaker when the

match was based on more precise information (suggested by the interaction terms). Depending on the information proxy, the point estimates are between 4 and 7 percentage points lower when there was more information, although the difference is not statistically significant when we use pre-hire employment as the information proxy (see column (1)). In sum, these findings suggest that matches formed with less precise information are more likely to be affected by revelations of worker-absence-type.

4 Conclusions

We document that workers matched to jobs with few internal substitutes are significantly less absent from work, compared to other workers in the same narrowly defined occupations. The difference is substantial and holds irrespectively if we look at employees' own sickness absence or absences among parents caused by child sickness. Parents working in jobs with lower employee substitutability shift part of their child sick absence spells to their partners.

About half of the correlation remains when we account for worker fixed effects, suggesting that both sorting based on pre-hire absence types and on-the-job changes in absence behavior are important mechanisms behind the strong association between sickness absence and employee substitutability. But sorting also occurs via the exit margin, as job separations respond to realizations of absence among new hires, particulary when they have few substitutes.

In addition, we find suggestive evidence that employee selection is more pronounced when there was more information about the workers' absence-types beforehand. Thus, screening leads to more efficient matching between workers of different absence types and jobs with different possibilities of internal replacement. Finally, we find that the separation response due to realized sickness absence among workers in jobs with few substitutes is negatively related to the amount of information in the hiring stage, suggesting that learning about match quality is an important determinant of turnover rates as in Jovanovic (1979).

Overall, our results highlight the importance of internal labor markets for

firms to handle the costs of production disruptions caused by work absence. For jobs with low internal substitutability, sickness absence is a significant determinant in the selection process of new workers. But the difficulties of perfectly predicting the absence propensity of new employees leads to mismatch between workers and firms and in turn job separations. Our findings thus validate previous theoretical and empirical work on the importance of sorting and point at sickness absence as a previously unexplored dimension of match quality.

From the worker's perspective, our findings suggest that episodes of sickness absence affect the individual chances of accessing and retaining unique positions, which account for a significant share of the labor market. Hence, workers have strong incentives to keep absence low in jobs with low internal substitutability, which they do by e.g. shifting child care to their partners. In future work it would be valuable to further explore if there is more scope for statistical discrimination against workers with above-average sickness absence rates at the group level in unique positions, e.g. women with children or workers from the upper part of the age distribution. Further explorations of the interplay between job characteristics and the allocation of time within the household could potentially also enhance our understanding of the systematic gender pay differences in modern labor markets.

References

Abowd, J. M., Haltiwanger, J., Lane, J., McKinney, K. L., and Sandusky, K. (2007). "Technology and the Demand for Skill:An Analysis of Within and Between Firm Differences." NBER Working Papers 13043, National Bureau of Economic Research, Inc.

Altonji, J. G. (2005). "Employer learning, statistical discrimination and occupational attainment." *The American economic review*, *95*(2), 112–117.

Andersson, F., Freedman, M., Haltiwanger, J., Lane, J., and Shaw, K. (2009). "Reaching for the stars: Who pays for talent in innovative industries?*." *The Economic Journal*, *119*(538), F308–F332.

Barmby, T., and Stephen, G. (2000). "Worker absenteeism: why firm size may matter." *The Manchester School*, *68*(5), 568–577.

Black, C. D., and Frost, D. (2011). *Health at work-an independent review of sickness absence*, vol. 8205. The Stationery Office.

Boye, K. (2015). "Care more, earn less? the association between care leave for sick children and wage among swedish parents." *IFAU Working Paper*, *2015*(18).

Dionne, G., and Dostie, B. (2007). "New evidence on the determinants of absenteeism using linked employer-employee data." *Industrial & Labor Relations Review*, *61*(1), 108–120.

Dustmann, C., Glitz, A., Schönberg, U., and Brücker, H. (2015). "Referral-based job search networks." *The Review of Economic Studies*, *forthcoming*.

Eriksson, S., Johansson, P., and Langenskiöld, S. (2012). "What is the right profile for getting a job? a stated choice experiment of the recruitment process." *IZA Working Paper*, *6691*.

Eriksson, S., and Rooth, D.-O. (2014). "Do employers use unemployment as a sorting criterion when hiring? evidence from a field experiment." *The American Economic Review*, *104*(3), 1014–1039.

Farber, H. S., and Gibbons, R. (1996). "Learning and wage dynamics." *The Quarterly Journal of Economics*, *111*(4), 1007–1047.

Fredriksson, P., Hensvik, L., and Nordström Skans, O. (2015). "Mismatch of talent: Evidence on match quality, job mobility and entry wages." *IFAU Working Paper*, *2015*(26).

German Federal Statistical Office (2011). *Statistical Yearbook 2011 for the Federal Republic of Germany*. Stuttgart: Metzler-Poeschel.

Goldin, C., and Katz, L. F. (forthcoming). "The most egalitarian of all professions: Pharmacy and the evolution of a family-friendly occupation." *Journal of Labor Economics*.

Hensvik, L., and Nordström Skans, O. (forthcoming). "Social networks, employee selection and labor market outcomes." *Journal of Labor Economics*.

Jackson, C. K. (2013). "Match quality, worker productivity, and worker mobility: Direct evidence from teachers." *Review of Economics and Statistics*, *95*(4), 1096–1116.

Jäger, S. (2015). "How substitutable are workers? evidence from worker deaths."

Jovanovic, B. (1979). "Job matching and the theory of turnover." *The Journal of Political Economy*, 972–990.

Kwon, I. (2005). "Threat of dismissal: incentive or sorting?" *Journal of Labor Economics*, *23*(4), 797–838.

Lange, F. (2007). "The speed of employer learning." *Journal of Labor Economics*, *25*(1), 1–35.

Lazear, E. P. (1998). *Hiring risky workers*. Springer.

Lazear, E. P. (2009). "Firm-specific human capital: A skill-weights approach." *Journal of Political Economy*, *117*(5), 914–940.

Lazear, E. P., and Oyer, P. (2007). "Personnel economics." *NBER Working Paper Series*, 13480.

Lindgren, K.-O. (2012). "Workplace size and sickness absence transitions." *IFAU Working Paper*, *2012*(26).

Mörk, E., Sjögren, A., and Svaleryd, H. (2013). "Cheaper child care, more children." *Journal of Population Economics*, *26*(1), 33–65.

Nagypál, É. (2007). "Learning by doing vs. learning about match quality: Can we tell them apart?" *The Review of Economic Studies*, *74*(2), 537–566.

Office for National Statistics (UK) (2014). *Full report: Sickness Absence in the Labour Market*.

Ose, S. O. (2005). "Working conditions, compensation and absenteeism." *Journal of health economics*, *24*(1), 161–188.

Oyer, P., Schaefer, S., et al. (2011). "Personnel economics: Hiring and incentives." *Handbook of Labor Economics*, *4*, 1769–1823.

Previa (2013). *Arbetsgivarnas kostnader för sjukfranvaro ökade med en miljard*. private company in the Swedish health industry.

Sattinger, M. (1975). "Comparative advantage and the distributions of earnings and abilities." *Econometrica: Journal of the Econometric Society*, 455–468.

Schönberg, U. (2007). "Testing for asymmetric employer learning." *Journal of Labor Economics*, *25*(4), 651–691.

Tinbergen, J. (1956). "On the theory of income distribution." *Weltwirtschaftliches archiv*, 155–175.

A Appendix material

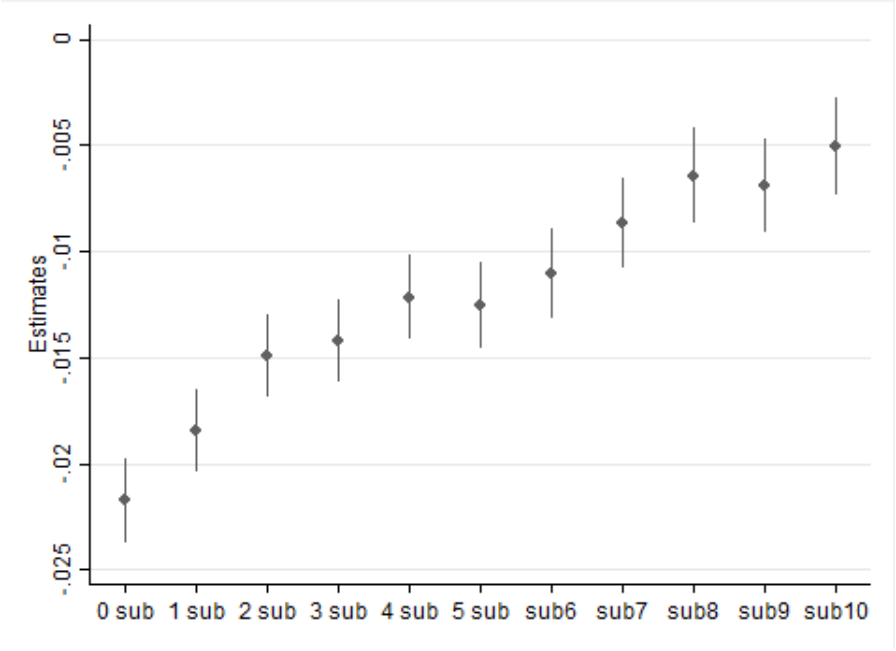

Figure A1: Sickness absence and the number of employee substitutes
Notes: The standard errors are clustered on the establishment level. The reference category is employees with more than 10 substitutes. The background controls are gender, age, education, birth country, having small children and establishment size. The model also includes year fixed effects, occupational fixed effects and establishment fixed effects.

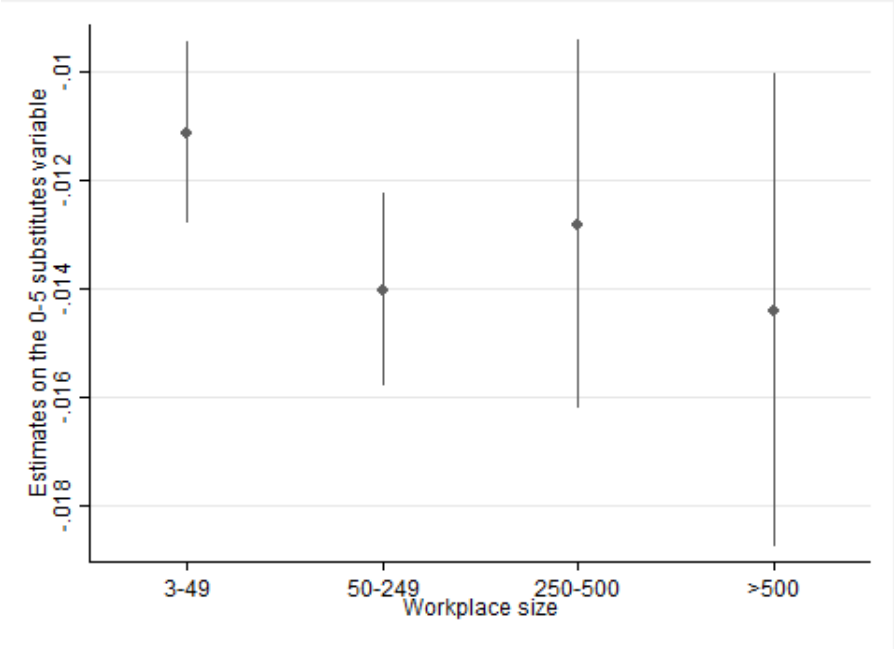

Figure A2: Sickness absence and workplace size

Notes: The figure shows the results from separate estimations of equation 1 by establishment size. The standard errors are clustered on the establishment level. The reference category is employees with more than 5 substitutes. The background controls are gender, age, education, birth country, having small children and establishment size. The model also includes year fixed effects, occupational fixed effects and establishment fixed effects.

Table A1: Descriptive statistics for all employees

	All	$0-5$ subst.	5+ subst.
Establishment characteristics			
No substitutes	0.036	0.204	0.000
Establishment size	570.9	90.8	675.7
Demographics			
Age	40.6	42.6	40.2
Male	0.629	0.495	0.658
Number of children aged 0-17	0.829	0.836	0.828
Country of origin			
Sweden	0.913	0.939	0.908
Rest of Nordic countries	0.032	0.026	0.033
Rest of Europe	0.026	0.018	0.027
North America	0.001	0.001	0.001
South America	0.006	0.003	0.007
Rest of the world	0.022	0.012	0.024
Education			
Pre high school education ($<$ 9 years)	0.025	0.021	0.026
Pre high school education (($\leq$ 9 years)	0.054	0.042	0.057
High school education max 2 years	0.413	0.386	0.418
High school education 23 years	0.225	0.233	0.223
Post high school education ($<$ 3 years)	0.143	0.169	0.138
Post high school education ($\leq$ 3 years)	0.131	0.143	0.128
Postgraduate education	0.008	0.004	0.008
Wage and Benefits			
Monthly wage in t (SEK)	23,657	22,824	23,839
Sickness benefit recipient in t	0.109	0.098	0.111
Professions			
Professionals	0.165	0.191	0.159
Technicians	0.245	0.315	0.229
Clerks	0.124	0.212	0.105
Service workers and shop sales	0.088	0.093	0.087
Skilled agricultural and fishery	0.005	0.010	0.004
Craft and related trades workers	0.116	0.090	0.121
Plant and machine operators	0.196	0.043	0.230
Elementary occupations	0.062	0.047	0.065
Number of observations	5,863,497	1,050,017	4,813,480

Notes. The sample is based on private sector employees in Sweden in 1997-2007. Managers and labor market entrants are excluded. The distribution across occupations is reported at the 1-digit level of the occupation code.

Table A2: Descriptive statistics for new hires

	All	$0-5$ subst.	5+ subst.
Establishment characteristics			
No substitutes[a)]	0.039	0.207	0.000
Establishment size	406.8	77.0	483.7
Demographics			
Age	35.7	37.4	35.3
Male	0.599	0.501	0.622
Number of children aged 0-17	0.807	0.890	0.787
Country of origin			
Sweden	0.898	0.928	0.891
Rest of Nordic countries	0.024	0.023	0.025
Rest of Europe	0.030	0.021	0.032
North America	0.001	0.001	0.001
South America	0.010	0.006	0.011
Rest of the world	0.065	0.041	0.071
Education			
Pre high school education (<9 years)	0.008	0.006	0.008
Pre high school education ((≤ 9 years)	0.089	0.059	0.097
High school education max 2 years	0.343	0.332	0.346
High school education 23 years	0.270	0.270	0.270
Post high school education (<3 years)	0.133	0.158	0.127
Post high school education (≤ 3 years)	0.150	0.172	0.145
Postgraduate education	0.006	0.004	0.006
Wage and Benefits			
Monthly wage in t (SEK)	22,413	22,226	22,457
Pre-hire sickness benefit recipient	0.116	0.125	0.114
Professions			
Professionals	0.171	0.186	0.168
Technicians	0.227	0.293	0.212
Clerks	0.128	0.207	0.110
Service workers and shop sales	0.133	0.113	0.137
Skilled agricultural and fishery	0.006	0.010	0.004
Craft and related trades workers	0.091	0.082	0.094
Plant and machine operators	0.165	0.052	0.192
Elementary occupations	0.079	0.058	0.084
Number of observations	387,901	73,366	314,535

Notes. The sample is based on private sector employees in Sweden in 1997-2007. Managers and labor market entrants are excluded. The distribution across occupations is reported at the 1-digit level of the occupation code.[a)] The fractions of new hires having 0 substitutes in our 8 broad occupation groups are: Professionals (5.3 percent), Technicians and associate professionals (4.7 percent), Clerks (8.2 percent), Service workers and shop sales workers (1.4 percent), Skilled agricultural and fishery workers (5.9 percent), Craft and related trades workers (2.7 percent), Plant and machine operators and assemblers (0.8 percent) and Elementary occupations (3.4 percent). This indicates that truly unique positions are present in all occupational skill levels.

Table A3: Descriptive statistics w.r.t. establishment size

(1) Est. size	(2) Share est.	(3) Share employees	(4) Share variation (0-5 subst.)	(5) Share variation (alt. def. I)	(6) Share variation (alt. def. II)
3-49	0.853	0.290	0.413	0.305	0.267
50-249	0.124	0.323	0.395	0.401	0.309
250-499	0.014	0.128	0.097	0.135	0.138
≥500	0.008	0.259	0.094	0.159	0.286

Notes. Columns (1) and (2) show the distribution of establishments and employees over establishment size. In columns (3)-(5) we show the share of the variation in having few employee substitutes, conditional on all covariates in equation (1) by establishment size. In column (3) we use our baseline definition (i.e. the number of coworkers in the same establishment and 3-digit occupation is 0-5). In column (4) we count workers as having few employee substitutes if they (i) have 0 substitutes in a workplace with 3-49 employees; (ii) have less than four substitutes in a workplace with 50-249 employees (iii) have less than seven substitutes in a workplace with 250-500 employees or (iv) have less than 10 substitutes in a workplace with more than 500 employees. In column (3) we divide the number of substitutes by the total number of employees on the workplace and require that quotient to be lower than 0.03 for an employee to be regarded as having few substitutes. Both these definitions imply that more of the identifying variation comes from larger establishments.

Table A4: Alternative definitions of employee substitutability

	(1)	(2)	(3)	(4)
	Outcome: Present sickness absence			
Definition of substitutability:	Baseline	Alt. def. I	Alt. def. II	
Low substitutability	-0.0131***	-0.0123***	-0.0131***	
	(0.0006)	(0.0006)	(0.0007)	
Share of substitutes				0.027***
				(0.0012)
Number of observations	5,863,497	5,863,497	5,863,497	5,863,497
Mean of dependent variable	0.109	0.109	0.109	0.109
Background controls	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes	Yes
Establishment fixed effects	Yes	Yes	Yes	Yes

Notes. The standard errors are clustered on the establishment level. See Table A3 for the alternative definitions of employee substitutability. The background controls are gender, age, education, birth country, having small children, and establishment size. $*$, $**$, and $***$ denote statistical significance at the 10-,5-, and 1-percent level.

Table A5: Substitutability and wages

	(1)	(2)
	Outcome: Log of monthly wage in t	
Low substitutability	0.0134***	0.0109***
	(0.0011)	(0.0010)
Number of observations	5,863,497	5,863,497
Mean of dependent variable	10.01	10.01
Background controls	No	Yes
Year fixed effects	Yes	Yes
Occupation fixed effects	Yes	Yes
Establishment fixed effects	Yes	Yes

Notes. The standard errors are clustered on the establishment level. The reference category is employees with more than 5 substitutes. The background controls are gender, age, education, birth country, having small children, pre-hire employment status and establishment size. $*$, $**$, and $***$ denote statistical significance at the 10-, 5-, and 1-percent level.

Table A6: Robustness checks

	(1)	(2)	(3)	(4)
	Outcome: Present sickness absence			
	Baseline	With wage control	Including public sector	4-digit occupation code
Low substitutability	-0.0131***	-0.0128***	-0.0079***	-0.0112***
	(0.0006)	(0.0006)	(0.0004)	(0.0010)
Number of observations	5,863,497	5,863,497	12,160,539	1,656,960
Mean of dependent variable	0.109	0.109	0.125	0.105
Background controls	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes	Yes
Establishment fixed effects	Yes	Yes	Yes	Yes

Notes. The standard errors are clustered on the establishment level. The reference category is employees with more than 5 substitutes. The background controls are gender, age, education, birth country, having small children and establishment size. Column (1) repeats the estimate from Table 1, col (2). In column (2) wage is included in the model. In column (3) we include public sector employees. In column (4) we calculate the number of substitutes based on a 4-digit occupational code which is available for the years 2005-2007. The occupational fixed effects are also based on the 4-digit occupational code. $*$, $**$, and $***$ denote statistical significance at the 10-, 5-, and 1-percent level.

Table A7: Absence on the extensive and intensive margin

	(1)	(2)
Margin:	*Extensive*	*Intensive*
Outcome:	Incidence (Baseline)	Log of benefits
	A. Present sickness absence	
Low substitutability	-0.0131***	-0.0195**
	(0.0006)	(0.0078)
Number of observations	5,863,497	638,409
Mean of dependent variable	0.109	4.822
	B: Own care leave	
Low substitutability	-0.0115***	-0.0381***
	(0.0017)	(0.0038)
Number of observations	1,911,734	1,249,558
Mean of dependent variable	0.654	3.605
	C: Partner's care leave	
Low substitutability	0.0087***	0.0143***
	(0.0018)	(0.0043)
Number of observations	1,767,118	977,270
Mean of dependent variable	0.553	3.427
Background controls	Yes	Yes
Year fixed effects	Yes	Yes
Occupation fixed effects	Yes	Yes
Establishment fixed effects	Yes	Yes

Notes. The standard errors are clustered on the establishment level. The reference category is employees with more than 5 substitutes. The background controls are gender, age, education, birth country, having small children and establishment size. The estimates in Column (2) are conditional on having positive benefits (i.e. at least one spell) according to the measure of interest. $*$, $**$, and $***$ denote statistical significance at the 10-, 5-, and 1-percent level.