



ELSEVIER

Contents lists available at ScienceDirect

## Journal of Public Economics

journal homepage: [www.elsevier.com/locate/jpube](http://www.elsevier.com/locate/jpube)Spillovers from gatekeeping – Peer effects in absenteeism<sup>☆</sup>Anna Godøy<sup>a, b, \*</sup>, Harald Dale-Olsen<sup>b</sup><sup>a</sup>Institute for Research on Labor and Employment, University of California, Berkeley, United States of America<sup>b</sup>Institute for Social Research, Oslo, Norway

## ARTICLE INFO

## Article history:

Received 20 December 2016

Received in revised form 23 August 2018

Accepted 24 August 2018

Available online xxxx

## JEL classification:

H55

I13

J22

Z13

## Keywords:

Sickness absence

Social interaction

General practitioners

Co-workers

Panel data

## ABSTRACT

We study peer effects in absenteeism among workplace colleagues. Gatekeeping is an essential task in many insurance systems. In this study we exploit exogenous shifts of general practitioners (GPs) occurring when physicians quit or retire. We find that these shifts induce changes in absenteeism for affected workers. By utilizing high-quality Norwegian matched employer–employee data with detailed individual information on certified sick leave during the period 2003–2012, we can study how the transfer of workers between GPs affects co-workers' absenteeism. We identify strong causal positive peer effects in absenteeism: a one day change in focal worker sickness absence transfers to a 0.41 day shift in peer absence.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The role of social preferences and norms in determining individual behavior and effort choices has been studied both theoretically and in lab experiments. A substantial literature has examined how colleagues influence each other through work ethics (Casadesu–Masanell, 2004), altruism and reciprocity (Adams and Rosenbaum, 1962; Akerlof, 1982) and fairness considerations (Adams, 1963; Akerlof and Yellen, 1990; Fehr and Schmidt, 1999). Individual worker behavior is typically not directly observable, making it difficult for firms to make contracts explicitly conditional on employee effort. Making contracts conditional on realized output will be costly for firms whose risk averse workers must be compensated for expected fluctuations in income. With incomplete contracts, social preferences

and norms can be an important factor in determining effort behavior at work (Fehr and Gächter, 2000).

In this paper, we analyze social spillover effects in worker absenteeism. Sickness absence is costly, both for business and public finances. In OECD countries, the cost of disability and sickness programs is much higher than spending on unemployment (OECD, 2010): In 2007, OECD countries spent an average of 0.8% of GDP on private and public sick leave programs alone. Moreover, the cost of absenteeism to firms may exceed the cost of sick pay due to disruptions to production. Peer effects in sickness absence may amplify such distortions.

Sickness absence is notoriously difficult for employers to control directly, as employee health is private information, observable only to the employees themselves and, to some degree, their physicians. In addition, the institutional context we study is such that workers have few economic incentives not to call in sick – during short term sickness absence, replacement ratios of benefits are high (typically 100%), and workers are legally covered by job protection legislation.

Economic theory suggests peer pressure may give rise to social spillover effects in absenteeism through local effort norms (Kandel and Lazear, 1992). Identifying such peer effects empirically is challenging however, as coworkers tend to be similar to each other at the outset. Moreover, coworkers may be subject to correlated shocks – e.g.

<sup>☆</sup> The research has received support from the Norwegian Research Council (grant number 227117). We thank three anonymous reviewers as well as conference participants at EEA-ESEM (Geneva 2016) and EALE (Ghent 2016) for their valuable comments. Data made available by Statistics Norway have been essential.

\* Corresponding author.

E-mail addresses: [anna.godoy@berkeley.edu](mailto:anna.godoy@berkeley.edu) (A. Godøy), [harald.dale-olsen@socialresearch.no](mailto:harald.dale-olsen@socialresearch.no) (H. Dale-Olsen).

similar work related health risks - that influence absence patterns, giving rise to a spurious within-group correlation in absenteeism.

To identify causal spillover effects, our empirical strategy focuses on the absence patterns of individuals whose colleagues experience an arguably exogenous shift in absence rates. In Norway, all residents are registered with a general practitioner (GP). These doctors act as the primary gatekeepers for paid sick leave, as all sickness absence lasting longer than 3 days must be certified by a physician. The basic premise of our identification strategy is that doctors will differ in their certification behavior, even when faced with identical patients.

When a GP quits or retires, their entire patient lists are typically sold along with the practice. As a consequence, an entire group of patients is shifted between two physicians with potentially different certification behavior. This allows us to compare the certification behavior of two doctors who face the same patients, recovering an unbiased measure of the difference between the two doctors' underlying certification propensities. We show that estimated certification propensities are significant in explaining changes in absence rates of the transferred patients.

Next, we use the estimated physician effects to estimate spillover effects on the focal workers' colleagues. As these colleagues are not directly affected by the physician transfer, any effect on this group can be interpreted as spillover effects. With this approach, we identify significant spillover effects in absenteeism among peers at work: depending on specification, a one percentage point increase in absence rate of focal workers increases the absence rates of similar age colleagues by up to 0.41 percentage points.

Estimated effects are stronger for coworkers who are close in age to the focal worker, which is in line with what we would expect if our estimates reflect social contagion. Extended models find that the effect is indeed behavioral and not driven by infectious diseases spreading among colleagues: focal worker absence increases peer absence that is due to non-communicable conditions (musculoskeletal, psychological). If anything, peer absence due to respiratory infections actually tends to fall slightly when the focal worker's absence increases, indicating that encouraging sick employees to stay home rather than go in to work may reduce the spread of contagious diseases at the workplace.

During the last decade, several studies have addressed social interaction issues related to sick leaves (Hesselius et al., 2009; Dale-Olsen et al., 2015; Lindbeck et al., 2016), disability receipt (Rege et al., 2012; Dahl et al., 2014a), welfare utilization (Åslund and Fredriksson, 2009; Markussen and Røed, 2015) and parental leave (Dahl et al., 2014b). These studies indicate a strong presence of social interaction effects.

Causal identification of peer effects using observational data is challenging (Manski, 1993): as individuals sort themselves into peer groups, outcomes tend to be correlated within peer groups even in the absence of causal peer effects. One identification strategy used to overcome these problems involves studying some reform or experiment which affected a group of individuals, identifying social interaction effects by measuring changes in outcomes among non-affected individuals. Several of the studies above follow such an approach directly, e.g., Hesselius et al. (2009) and Dahl et al. (2014b), while others achieve this indirectly (e.g., Dale-Olsen et al. (2015) exploited a tax reform which affected a limited number of workers).

Hesselius et al. (2009) was the first study to convincingly identify peer effects in sick leave behavior among colleagues. The authors utilized variation from a 1988 experiment in Gothenburg, Sweden, where half the city's population were randomly assigned treatment in the form of increased maximum duration of self-certified sick leave (12 days for the treated versus 6 days for the control group) – the experiment significantly increased absence rates in the treated group. Hesselius and co-authors find that as the share of treated workers at the workplace increases, so do the sick leave days of the untreated

colleagues, i.e., the untreated workers respond to the behavior of their colleagues.

While our paper is clearly related to Hesselius et al. (2009), the value-added is considerable. Our key result is that we show that the peer effects in sick leave behavior is not limited to self-certified absence from work, but even extend to physician-certified sick leaves. This is a relevant finding in its own right, as it indicates that the presence of gatekeepers does not stop these peer effects from happening.

From an economic policy point of view this is important, as physician-certified sick leave tends to have a greater public finance effects compared to short term absence. Physician-certified sick leaves constitute the majority of the lost work days in most countries.<sup>1</sup> The distribution of sickness absence is highly skewed, with long term absence accounting for most of the cost of sick pay. Moreover, many welfare regimes follow a pattern where sick pay for short term, self-certified absences are covered by the employer or not at all, while long term, physician-certified absences are covered by public authorities.

The findings in this paper point to a policy lever to reduce long term absence rate. Our research design highlights the importance of the GP as a gatekeeper in the welfare system, while pointing out likely multiplier effects. In the presence of moral hazard, policymakers will often face a difficult tradeoff between providing full insurance – 100% sick pay – and maintaining incentives for work and economic self-sufficiency. Stricter gatekeeping is often proposed as a way to reconcile these two policy objectives. Our findings predict that increased gatekeeping will have multiplier effects, effectively magnifying the impact of these policies: in our policy simulations, we find that spillover effects account for 43% of the impacts of a simulated gatekeeping reform.

Finally, our data includes information on diagnosis-specific absence rates, allowing us to examine the pattern of peer effects in more detail. Specifically, we show how the peer effects arise through specific complaints and disorders, and can discuss our results in relation to transmittable diseases, effort-provision and workload.

A number of papers have used variation in gatekeeper leniency to identify effects of disability insurance (DI), by exploiting random assignment of adjudicators and medical examiners (Maestas et al., 2013; French and Song, 2014). One particularly relevant paper is Dahl et al. (2014a)'s paper analyzing the intergenerational transmission of disability insurance (DI) enrollment. Applicants who were assigned more lenient judges were more likely to be granted DI on appeal. In a second step, applicants were matched to their adult children, using the random variation in judge leniency to identify "family welfare cultures". The authors found evidence of significant spillover effects: persons whose parents were assigned more lenient judges were themselves more likely to be enrolled in DI as adults.

Our paper differs from this literature in one important aspect: in the case of GPs, gatekeepers are not randomly assigned. In particular, patients may self-select to more lenient GPs in order to get more absence days. Our identification strategy then relies not on random assignment of the initial GP, rather we argue that the change in GP induces a shift in certified absence patterns that is as good as random. These GP changes have been used as a natural experiment to estimate the impact of GPs on absenteeism (Markussen et al., 2013). We discuss this assumption of random assignment in more detail in Section 3.

The paper whose empirical approach perhaps most closely resembles that of the present paper is Dahl et al. (2014b)'s work on peer effects in the take-up of paternity leave. In the paper, the

<sup>1</sup> While the numbers vary over the period we study, on average 10–20% of absence in Norway is self-certified, with the rest being physician-certified.

authors exploit a reform in the Norwegian paid parental leave policy. In 1993, the system was changed to encourage fathers to take more parental leave: for children born after April 1, 1993, 4 weeks of the total paid parental leave period were reserved for fathers. Using a regression discontinuity design comparing fathers of children born just before and just after the cut-off date, the authors found that this reform significantly increased take-up rates of the affected fathers. In a second step, the authors identified significant peer effects: co-workers and brothers of men who had children just after the reform were significantly more likely to take parental leave compared to peers of fathers who had children just before the reform was implemented. While different thematically, the identification strategy of their paper is fundamentally similar to ours: A clearly defined subset of individuals were quasi-randomly assigned to treatment; this source of variation is used to identify spillover effects in a linked sample of non-affected peers.

The rest of the paper is organized as follows: Section 2 gives a brief overview of the relevant institutional context. Section 3 presents our empirical strategy and econometric models. Section 4 describes the data together with some summary statistics. Our findings are presented in Section 5. Section 6 shows results from extended models that examine potential mechanisms. Section 7 concludes.

## 2. The Norwegian sick pay legislation and the role of the GP

The public sick pay system in Norway is prototypical of a generous Scandinavian welfare state. Coverage is universal, for both full time and part time employees, and the replacement rate is generally 100%.<sup>2</sup> Individuals can receive sick pay for a period of up to 12 months. After benefit exhaustion, persons who are unable to return to work are transferred to significantly less generous temporary disability insurance benefits.

While short absence spells lasting up to 3 days can be self-certified, most sickness absence must be certified by a physician. This certification might be provided by a hospital, a private sector doctor, or a patient's GP. Most absence certificates are issued by GPs. The practice of GPs to certify a sick leave is found in most countries (OECD, 2010).

During the period studied in this paper (2003–2012), the average absence rate was 7.0 %, including both self-certified (0.9%) and physician-certified (6.1%) absence.<sup>3</sup> There are substantial differences between different industries: while the average absence rate in the healthcare sector was 9.6%, people working in technical and professional services were absent just 4.7% of the time.

Since 2001, all Norwegian residents have been registered with a GP. These GPs are primarily publicly funded.<sup>4</sup> GPs are typically self-employed in private practice, and are compensated on a two tier schedule: They receive a fixed payment according to the number of patients on their lists, together with payments for services rendered (consultations and procedures).<sup>5</sup>

Initial assignment of patients to GPs is based on their residential address. However, individuals are free to choose any doctor they like as their GP, subject to capacity constraints, i.e. if the doctor

has any vacant slots on their lists. Assignment of patients to doctor is thus typically not random. In particular, patients who choose to change doctors may be motivated to do so in part because they are not satisfied with their previous doctor's certification practice. Markussen and Røed (2017) find indications that this is indeed the case: doctors who have a history of being lenient in issuing absence certificates face higher demand from patients seeking to change their GP assignment, compared to their stricter colleagues.

Meanwhile, when GPs retire, close their business for one reason or another (e.g. moves to another part of the country), their entire list is typically sold to a new GP together with the rest of the practice. This creates a potentially exogenous shift in the relationship between patients and GPs.

The transfer process begins with the GP sending the municipality their resignation. The resignation should be given with 6 months' notice. The municipality then posts the vacancy – the vacancy must be posted in the journal of the Norwegian medical association as well as in at least one national newspaper, and it must include relevant information on the practice, including the number of patients on the list and whether it is a group practice. The municipality then advises all applicants on the projected price of the practice. The municipality is also responsible for reviewing and selecting candidates. For our research design, it is important to note here that the incumbent GP plays no formal role in the selection process.

The actual sale of the practice thus only happens after the replacement GP has been selected by the municipality. The sale includes the office and medical equipment, as well as the value of the patient list. The price is typically arrived at through negotiations; if this fails, the price is set by binding arbitration. The replacement GP then takes over the care of patients at the agreed upon date, typically 6 months after the resignation was received by the municipality.

The role of the GPs and family doctors and their importance are discussed and analyzed in several theoretical and empirical studies (Blomqvist, 1991; Grytten and Sørensen, 2003; Dusheiko et al., 2006; Markussen et al., 2011, 2013). Primary physicians' primary goal is to treat illness in all its shapes and forms. However, at the same time GPs act as gatekeepers for the welfare state: they are to certify that an individual really is ill and thus is eligible for sick pay (insurance). This creates a tension (Blomqvist, 1991). Several studies have identified variation in the practice of doctors, GPs and family doctors, from hospital admission rates (Dusheiko et al., 2006), to the use of expensive laboratory tests (Grytten and Sørensen, 2003), to sick leave behavior in general (Markussen et al., 2011, 2013). Markussen et al. (2013) is particularly relevant in our case, since they, like the present paper, analyze the Norwegian system of GPs using transfers of patient lists from doctors who exit the market. Markussen et al. (2013) conclude that "The key finding of our paper is that family doctors really do have a significant impact on their patients' benefit claims".

## 3. Empirical strategy

The goal of this paper is to identify spillover effects in absenteeism among peers at work. Let  $y_{it}$  denote the absence rate of individual  $i$  in year  $t$ . A naive model of spillover effects can be formulated as

$$y_{it} = x_{it}\beta + \bar{y}_{j(i)t}\tilde{\gamma} + \varepsilon_{it} \quad (1)$$

where  $\bar{y}_{j(i)t}$  denotes the average absence rate in  $i$ 's peer group  $j$ .

Estimating Eq. (1) by simple OLS may reveal correlations between own absence and average peer absence rates. However, this correlation should be interpreted with caution, and should not be given a

<sup>2</sup> Formally, benefits are capped for high earnings; in 2015, the threshold was 540,408 NOK or 57,480 Euro. However, all public sector workers and many private sector workers are covered by additional top-up insurance, ensuring 100% replacement ratio even for high incomes.

<sup>3</sup> Figures from Statistics Norway.

<sup>4</sup> There are copayments for patients aged 18 and over, however they are generally small and capped at around 2000 NOK (200 Euro).

<sup>5</sup> The law regulating the GPs allows for physician to work as salaried employees in exceptional cases. This is the case for a small minority of GPs. For these doctors the transfer process is largely similar to the one described here, with the exception of the sale of private practice between the two doctors.

causal interpretation. The empirical analysis has to overcome several barriers to identification (Manski, 1993).

First, the observed colleague peer groups are not randomly assigned. Rather, colleagues in the same firm are likely to be similar in terms of underlying characteristics that influence absence patterns, that may be only partially captured by  $x_{it}$ . Colleagues at the same firm may experience the same shocks to sickness absence – whether physical (cold and flu) or psychosocial (conflicts at work or other correlated shocks). Finally, if  $i$ 's absence is influenced by their colleagues, we would expect the reverse to be true –  $i$ 's absence patterns would influence  $\tilde{y}_{j(i)t}$ . This would imply that  $y_{it}$  appears on both sides of Eq. (1) (the reflection problem).

Our identification strategy focuses on cases where a single individual among  $i$ 's colleagues experiences an exogenous shock to their sickness absence. Letting  $y_{it}$  denote the absence rates of individual  $i$  at time  $t$ , and let  $y_{j(i)t}^*$  denote the absence rate of the focal worker who is their colleague in peer group  $j$ .

$$y_{it} = \theta_j + \theta_t + x_{it}\beta + y_{j(i)t}^*\gamma + \varepsilon_{it} \quad (2)$$

where  $\theta_j$  are peer group fixed effects,  $\theta_t$  are calendar time effects, and  $x_{it}$  is a vector of control variables.

Identification of Eq. (2) faces the same difficulties as outlined above. To identify  $\gamma$ , we implement a two-step approach. First, we construct an instrumental variable using individual data on certified sick leave linked to physicians. In the second step, we use these estimated physician effects to estimate models of spillovers in sickness absence. In the following, we will describe the construction of the instrument as well as our econometric models in more detail.

### 3.1. Identifying assumptions

In order to estimate causal effects of changes in GP leniency on coworker absence rates, the independence assumption must hold. This assumption states that conditional on the covariates of the model, variation in GP practice style must be uncorrelated with other confounding factors that may influence absence rates, such as coworker health or socioeconomic characteristics. When estimating IV models, additional assumptions are required: the exclusion restriction, relevance (existence of a first stage), and monotonicity. In this section, we discuss each of these assumptions in some detail.

**Independence:** The random assignment assumption states that the GP leniency should be independent on underlying absence propensities, conditional on the other variables in the model. As the models include fixed effects for peer group, this assumption needs to hold only within peer group. Note that we do not require that the initial GP assignment is random. On the contrary, we recognize that patients may have self selected to GPs in part due to their certification practice. However, we require that the change in GP practice should be exogenous to individual absence propensities.

We argue that the institutional context indicates that this assumption is likely to hold. The replacement GP is chosen by the municipalities, leaving little scope for the incumbent GP to influence the selection process. However, even if the GP change is exogenous to patient health, random assignment may be threatened if the change in GP is foreseen by patients. Patients may also choose to exit the list after the new GP takes over the practice if they do not like him or her, this decision may be influenced by certification behavior along with other aspects of practice style. As a consequence, the mix of patients who stay with the practice before and after the change may be endogenous to GP leniency.

To address this, we lock patients to the GP at the beginning of the calendar year prior to the transfer event, that is, we retain all patients who were registered with the incumbent GP in January the year before the transfer, regardless of whether they follow along to

the new GP or select away to a different GP. This effectively ensures that patients are locked to GPs at least 12 months before the transfer happens. As regulations require GPs to resign with 6 months notice, this restriction ensures that matches are defined before formal notice is given, reducing the scope for patients to select away from exiting GPs – patient would effectively have to be able to anticipate GP exits at least 6 months before the GP gives notice. To examine whether this is the case, we estimate a set of robustness exercises locking patients to GPs even earlier – 2 and 5 years before the event – results from this exercise are discussed in Section 5.

Next, we consider the implications for random assignment in the case where high propensity workers could be seeking out lenient GPs. When these lenient GPs exit, their high propensity patients will then, on average, match with GPs that are less lenient in comparison, creating a negative correlation between absence propensity and  $\Delta FE$ . Similarly, selection patterns in the physician labor market could also threaten random assignment. If lenient GPs tend to retain and attract patients with high absence propensity, they may build larger patient lists over time. If longer patient lists are seen as a positive in the GP job market, exiting lenient GPs may attract more candidates. If, in turn, municipalities prefer to hire stricter GPs, this pattern would induce a similar negative correlation between absence propensity and  $\Delta FE$ .<sup>6</sup> This kind of selection need not threaten random assignment – as long as this selection pertains to fixed characteristics of the focal worker, the model controls for this through the inclusion of peer group fixed effects.

We can get an indication of whether the independence assumption is likely to hold by examining correlations between the instrument and observable characteristics of peer and focal workers. For this exercise, we would ideally want to observe direct measures of worker health, either objective measures (blood pressure, body mass index) or subjective measures (self-perceived health). Though we do not observe these measures in our dataset, we have access to a rich set of demographic characteristics, including age, gender, education and work experience, that are significant in predicting absence rates. Using these demographic characteristics, we construct a measure of absence propensity, which we in turn link to the estimated change in GP effect. If random assignment holds, these characteristics should be uncorrelated with the change in GP effect  $\Delta FE$ . These exercises are presented in Section 5.

**Exclusion:** The exclusion restriction states that the instrument (GP fixed effects) should influence peer absence rates only through their effects on absence rates of the focal workers. Throughout the paper, we refer to the estimated doctor fixed effects as “leniency indicators”. However, GPs may vary across other dimensions than leniency, notably quality of care, affecting patients' health and in turn, their absence rates.

To illustrate, some doctors with a high value of  $\hat{\alpha}$  may provide lower quality care, reducing the health of the focal worker, increasing their absence rates even if there is no change in leniency per se. This need not violate the exclusion restriction: as long as the only channel influencing peer behavior is through the absence pattern of the focal workers, it does not matter why the focal worker's absence is increasing, in fact this may not even be observable to the peers.

However, the exclusion restriction will be violated if the change in focal workers health affects coworker health independently of focal worker absence rates. One potential channel for this is through physical contagion effects: if more capable GPs are able to treat and prevent diseases that are likely to spread at work, this would improve the health of both peer and focal workers. This could also be the case if there are information spillovers in health: the new GP may educate the focal worker on how to manage their health, and the focal worker could share this information with coworkers, reducing

<sup>6</sup> Assuming GP leniency is at least partially observable to the hiring municipality.

their absence rates independently of any change in absence patterns. This would violate the exclusion restriction.

Since we do not observe health status in our data, we cannot test for this directly. Still, there are testable implications. First, we can see how estimated effects vary by diagnosis. If the spillovers represent physical contagion, we would expect the effects to be driven by respiratory infections that are likely to spread among coworkers, rather than non-communicable conditions like musculoskeletal and psychological conditions. Second, we estimate augmented models where physician age and gender are included as additional control variables, acting as proxies for other aspects of practice style. If the estimates are sensitive to the inclusion of these controls, it could indicate that the exclusion restriction is threatened.

*First stage:* The first stage (relevance) assumption states that the excluded instrument - the estimated physician effects - should be significant in explaining focal worker absence rates. This assumption is testable directly by observing the F-statistic of the excluded instrument. We report this statistic wherever we report IV estimates.

*Monotonicity:* The monotonicity assumption states that the instrument should affect the endogenous variable being instrumented in the same direction for all focal workers who are affected by the instrument. That is, no individual should reduce their absence rate as a result of being transferred to a more lenient doctor - as measured by a higher estimated GP fixed effect.

The monotonicity assumption may be violated if GPs differ in their leniency between different patient groups. A GP may be more lenient toward a large majority of their patients, while being less lenient on a small subset (i.e. certain diagnoses or patient demographic characteristics). In this case, some patients may see their absence rates go down even as they are transferred to a GP with a higher estimated fixed effect, and the monotonicity assumption will be violated.

A testable implication is that the first stage should be non-negative for all subsamples. We discuss these models further in Section 5.

### 3.2. Constructing the instrument

The instrumental variable is obtained by estimating pairs of physician fixed effects using a sample of individuals who experience an exogenous transfer of physicians when their old doctor quits or retires. Unfortunately, the data does not include a variable indicating that such a transfer occurs. Instead, we use observed flows of patients between physicians to infer when a transfer has occurred.

Specifically, we define an exogenous transfer between two physicians ( $p$  and  $p'$ ) as occurring in year  $t$  if at least 85% of the patients registered with physician  $p$ 's patients are registered with physician  $p'$  in year  $t + 1$ .<sup>7</sup> That is, an individual is included in the sample used for estimating the GP effects if he or she was registered with physician  $p$  in year  $t - 1$ , regardless of whether they actually experienced the transfer. Let  $\mathcal{P}$  denote the set of patients registered to physician  $p$  in year  $t - 1$ , and let  $N_{ip}$  and  $N_p$  denote the number of observations of patient  $i$  at physician  $p$  and the total number of patient-year observations of physician  $p$ , respectively.

We estimate the following simple linear model of absence rates

$$y_{it} = \theta_i + x_{it}\beta + \varepsilon_{it} \quad (3)$$

where  $\theta_i$  are patient fixed effects and  $x_{it}$  is a vector of age dummies. Note that the inclusion of patient fixed effects implies the model controls perfectly for time invariant patient determinants of absence rates, both observable characteristics (e.g. gender) and unobservable

characteristics such as any health conditions that are stable over time.<sup>8</sup>

Eq. (3) is then estimated on a panel containing 4 years of absence data for each individual: from year  $t - 2$  to  $t + 2$ , excluding the year of the transfer. The residuals from this regression are then used to construct the instrument.<sup>9</sup>

For each patient  $i$  at GP  $p$ , the instrument is defined as the leave-out mean residual sickness absence of that GP's other patients:

$$\alpha_{ip} = \frac{1}{N_p - N_{ip}} \sum_{j \in \mathcal{P}, j \neq i} \hat{\varepsilon}_{jt}$$

That is, the constructed instruments are individual-specific. Calculating the leave-out mean in this manner avoids changes in individual  $i$ 's own absence rates correlating with the instrument.<sup>10</sup>

One concern when estimating models like Eq. (3) is that the estimated physician effects will suffer from omitted variable bias. The leniency indicators  $\alpha_p$  will typically capture not only the influence of doctor  $p$ , but also time-varying characteristics of doctor  $p$ 's patients, such as their underlying health status. However, our empirical strategy involves a pairwise comparison of estimated  $\alpha$ 's of two doctors who are facing the same patient group. As a consequence, any omitted variable bias should in expectation be the same for the two doctors. As a consequence, the difference  $\alpha_p - \alpha_{p'}$  is an unbiased estimate of the difference between the two physicians practice styles.

The estimated  $\alpha$  can be interpreted as a measure of gatekeeping, i.e. the leniency of the physician when faced with a request for a sickness absence certification. However, they can also capture other aspects of practice style, such as medical skill or skills at communicating with employers - patients of highly skilled doctors could, on average, require fewer and shorter absence spells. In addition, the transfer may in itself influence the doctor effect (Markussen et al., 2013): The new doctor will be unfamiliar with the patients, they may face stronger competitive pressure to comply with patients' wishes to retain them as clients, and the change of doctor may in itself lead to disruption in treatment or return-to-work efforts.

To summarize, the change in  $\alpha$  may reflect other aspects of physician practice style in addition to their willingness to comply when patients request absence certificates. However, it is important to note that this is not a threat to our identification strategy, as all these alternative channels work only through the treated individual's absence patterns.

### 3.3. Event-study models

We formulate an event-study model to see how absence patterns of focal workers and their colleagues change around the time of the

<sup>8</sup> The model does not include calendar time controls. As age and calendar time are perfectly collinear in the model with individual fixed effects, we cannot control for both without imposing additional parametric assumptions. An alternative specification that omits the individual fixed effects, instead including calendar time, age controls, gender and education yields very similar estimates.

<sup>9</sup> Note that patients that are involved in the transfer are assigned the same replacement physician, even if they opted out and chose a different physician.

<sup>10</sup> Although the "negative correlation bias" associated with the leave-out-mean approach (Angrist, 2014) could potentially increase the IV estimate, this bias should be exceedingly small given that the size of the GP patient panels averages 544 patients, and thus have only a trivial biasing effect. In a similar setting, discussing this negative correlation bias arising when estimating the effect of being assigned to "higher quality" classroom, Chetty et al. (2011) still end up focusing on the leave-out-mean instrument, even if class groups are many times smaller than GP-patient lists.

<sup>7</sup> Constructing the auxiliary and estimation samples using monthly data yields very similar results, and the gain in precision is negligible.

physician transfers.<sup>11</sup> Specifically, we examine how the change in absence rates changes relative to the change in physician FE.

Formally, we define  $\pi_{k(j,t)}$  as a set of indicator variables of relative time:

$$\pi_{k(j,t)} = 1(t - t_j^* = k)$$

where  $t_j^*$  denotes the year of GP transfer, and let

$$\delta_j = \hat{\alpha}_{p'j} - \hat{\alpha}_{pj}$$

denote the difference between the estimated FE of the new and the old doctor of the treated individual in peer group  $j$ .

We model annual absence rates of individual  $i$  in peer group  $j$ :

$$y_{it} = \theta_j + \theta_t + x_{it}\beta + \sum_{k=-4}^4 (\pi_{k(i,t)} \times \delta_j(i)) \gamma^k + \varepsilon_{it} \quad (4)$$

Here, the vector of control variables  $x_{it}$  includes the main effects of event time, allowing the physician change itself to impact absence rate, independent of change in practice style. The main coefficients of interest are the  $\gamma^k$ , which capture effects of the change in physician FE interacted with time since doctor change. These are normalized relative to the effect in the year before the physician change (year -1), i.e. we set  $\gamma^{(-1)} = 0$ .

For focal workers, we expect average absence rates to change nearly one to one with the change in estimated physician FE after the transfer. Notice that the event-study models are estimated on a panel that includes data up to 4 years before and 4 years after the physician transfer. As a result, estimating the model (4) on focal workers can serve as a check of whether estimated physician effects  $\hat{\alpha}$  reflect persistent physician behavior, rather than trends. If the former is the case, we would expect estimated  $\gamma$ s to be flat in the 4 years before and after the switch.

In the absence of spillover effects, we would expect no effects on the non-treated colleagues. If there are spillover effects, we would expect absence rates to increase in proportion with the increase in physician FE, though less than one-to-one.

### 3.4. Instrumental variable models

The estimated GP effects are then used to obtain IV-estimates of model (2), reproduced below:

$$y_{it} = \theta_j + \theta_t + x_{it}\beta + y_{j(i)t}^* \gamma + \varepsilon_{it}$$

Where the first stage can be written

$$y_{jt}^* = \theta_j + \theta_t + x_{it}\beta + \hat{\alpha}_{p(j,t)} \gamma^F + \varepsilon_{it} \quad (5)$$

The corresponding reduced form model linking peer absence to GP effects can be written:

$$y_{it} = \theta_j + \theta_t + x_{it}\beta + \hat{\alpha}_{p(j,t)} \gamma^{RF} + \varepsilon_{it} \quad (6)$$

As before,  $y_{jt}^*$  is certified sickness absence of the focal worker in peer group  $j$ , and  $\theta_j$  is a peer group fixed effect.

The sample will be constructed so that each peer group  $j$  is merged with exactly one focal worker. The parameter of interest  $\gamma$  will be identified from variation in the physician FE of the affected

worker, which occurs after the first doctor transfers their list to their successor. In other words, Eq. (2) will be identified from the same variation that is captured by the event-study models.

Recall that in the year of transfer, the focal worker receives care from both the incumbent and replacement GP. For this reason, the transfer year is excluded from the estimation sample when estimating the reduced form and instrumental variable models.

As discussed above, random assignment alone is sufficient for the reduced form model represented in Eq. (6) requires only random assignment in order to produce unbiased estimates of spillover effects of GP practice style on peer absence. In order for the instrumental variable approach to be valid, three additional conditions are required to hold: the exclusion restriction, relevance and monotonicity. Under these assumptions, the estimated effects can be interpreted as the local effects on the subpopulation of compliers - those individuals whose focal worker colleagues change their absence patterns as a result of being assigned a new physician.

## 4. Data

### 4.1. Sample

Sample construction proceeds in two steps. First, the sample of individuals experiencing an exogenous physician transfer is constructed; this is then used to estimate the instrumental variable (physician effects). In a second step, we attach data on co-workers of the transferred individuals, making up the main estimation sample.

The starting point of the sample is all individuals who experienced a mass transfer of GP during the years 2005–2010. Again, we define a transfer as occurring in year  $t$  if at least 85% of patients who are registered with doctor  $p$  at the start of year  $t$ , are registered with another doctor,  $p'$ , in year  $t + 1$ .<sup>12</sup> Typically, some patients will not comply with the transfer, instead choosing to pick a different doctor when their old doctor retires. Moreover, some patients may anticipate their doctor's retirement, and choose to seek out a new doctor before year  $t$ . As a way to address both these issues, we retain in our sample all patients registered with doctor  $p$  in year  $t - 1$ ; moves to doctors other than  $p'$  are ignored (treated as transfer to  $p'$ ).<sup>13</sup> Persons who experience more than one such transfer are excluded from the sample. This approach leads to a total of 409 physicians signing over their lists, with a sample average list size of 544 patients.

This sample is then merged to data on certified sickness absence for the years from  $t - 2$  to  $t + 2$ . The data include all certified absence, regardless of length of spell. Our sample thus includes both sick leave covered by the employer (typically the first 16 days) and government covered sick leave, however we do not observe self-certified absence. For each individual then, we have 2 observations of sickness absence with the exiting doctor  $p$ , and 2 observations with the new doctor,  $p'$ . Finally demographics data including age, gender, immigration background, as well as work situation are attached to the sample.

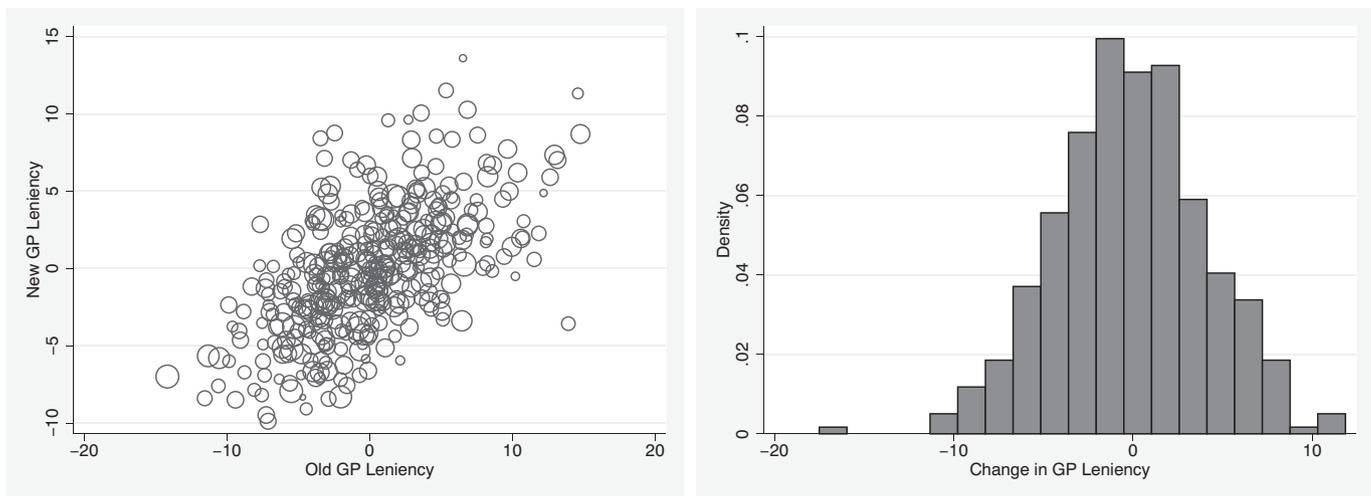
We then estimate Eq. (3) on the full sample, as well as separately by even/odd birth year. As discussed in the previous section, the estimated physician fixed effects will be biased (OVB), but the difference between pairs of physician FE will be a consistent measure of the difference in the practice style of the two doctors.

Fig. 1 illustrates the pattern of GP fixed effects before and after the transfer. The panel on the left plots the estimated GP effects before

<sup>12</sup> In Section 5, we discuss robustness to alternative cutoffs.

<sup>13</sup> By defining the transfer sample with respect to GP at the start of year  $t - 1$ , we retain patients even if they anticipate the GP switch with up to 12–23 months notice (depending on month of transfer). Given that the formal term of notice for exiting GPs is 6 months, this should limit the scope for self selection. Results are qualitatively robust to this timing, see Appendix, Table A5.

<sup>11</sup> Following the approach of Finkelstein et al. (2016).



**Fig. 1.** Estimated GP effects before and after transfer.

Note: Panel on the left plots estimated GP fixed effects for the replacement and exiting GPs. Panel on the right plots within doctor-pair differences in estimated doctor fixed effects from Eq. (3).

and after the change, while the panel on the right plots the estimated differences  $\delta_i = \alpha_{p'} - \alpha_p$ .<sup>14</sup>

The panel on the left indicates a positive correlation between the estimated GP effects before and after the change. The distribution of the estimated  $\delta$  appears to be roughly symmetrical around zero. On average, new GPs tend to have slightly higher absence propensities than the doctors they replace: 42% of transfers involve a move to a stricter doctor ( $\delta_i < 0$ ), while 58% of transfers involve a move to a more lenient GP ( $\delta_i > 0$ ).

Having estimated the physician FE, the next step is to construct the main estimation sample. The treated individuals in the auxiliary sample are then linked to groups of co-workers in the same firm. One limitation of the data is that we cannot observe directly the extent to which colleagues interact with each other. Our approach relies on linking the focal workers to colleagues who are around the same age, working in similar occupations. Specifically, the focal workers are linked to workers in the same establishment (in the case of multi-establishment firms), who are in the same age range and employed in the same occupation as the focal worker. The preferred estimation sample uses a  $\pm 2$  years age bracket, defining occupations on a three digit level, however alternative samples are constructed to explore consequences of increasing the allowable age difference and including peer workers in more broadly defined occupational categories.<sup>15</sup>

Age-occupation peer groups that are matched with more than one focal worker are excluded from the sample. Moreover, to increase the likelihood that the constructed peer groups do in fact reflect places of social interactions, very large peer groups ( $\geq 20$  peer workers) are excluded from the sample. This excludes 1.7% focal worker years from the sample; 0.7% of focal workers are excluded entirely.<sup>16</sup>

The sample is restricted to people who are employed in the same firm the full calendar year: persons who have more than one employer are excluded from the sample. Data is included for up to 4 years before and after the transfer year (i.e. maximum 8 observations per individual). When a focal worker moves out of the

firm-occupation group they held in year  $t^*$ , that focal worker and their peers are no longer included in the sample.<sup>17</sup>

Table 1 shows some summary statistics of the sample. The sample covers a total of 137,303 persons: 22,632 transferred patients and 114,671 workplace peers. On average, each focal worker is matched to 3 peers per year. Meanwhile, the average peer worker is in a group of 6 (non-treated) colleagues.

Treated and peer workers are fairly close in age; this follows from how the sample is constructed. Looking at Table 1, focal workers and matched peers are also similar in terms of gender and absence patterns, which is reassuring.

An important question in all empirical research is whether the findings are likely to generalize. To shed light on this question, we first see how the estimation sample compares to the full working population in terms of observable characteristics and absence patterns. As a consequence of how the peer groups are defined, employees of both very small and very large establishments are less likely to be included in the sample. The requirement that workers should be employed all year, by a single employer, means that we may be less likely to retain people who have a strong connection to the labor market. When comparing the estimation sample to a reference sample, the estimation sample is more male (56 vs 51%), older (mean age 44 vs 40) and have higher earnings.<sup>18</sup> Meanwhile, the estimation sample appears to be representative of the population in terms of education background and occupations of workers as well as absence patterns.

The instrumental variable model estimates local effects on compliers - the subset of patients who are induced to change their absence pattern as a result of the GP change. When evaluating external validity, we also need to take into account how these compliers are characterized relative to the sample population. We investigate this later following the approach of Imbens and Rubin (1997).

To get a first impression of the correlation between absenteeism in work peer groups, Fig. 2 shows a binned scatter plot of own absence days  $y_{it}$  and leave-out mean absence rates of the peers,  $\bar{y}_{j(i,t)}$ .

<sup>14</sup> Transfers involving  $< 50$  patients are excluded when plotting the figure, leaving out 0.06% of the auxiliary sample.

<sup>15</sup> Occupational groups are primarily defined using Statistics Norway's standard classification of occupation, which is based on EU's standard classification ISCO-88(COM).

<sup>16</sup> When presenting the results of the model, we discuss the robustness of our findings to removing this restriction.

<sup>17</sup> In the robustness section, we consider an alternative specification locking peer groups at year  $t^* - 1$ , ignoring changes in the focal worker's firm-occupation status in the eight year window around the transfer year. Results are qualitatively robust to this specification.

<sup>18</sup> The reference sample is a five percent random sample of the population aged 20–60, who were registered working at any point each year.

**Table 1**  
Summary statistics.

	(1)	(2)
	Peers mean	Treated mean
Age	44.4	44.3
Female	0.44	0.43
Years in sample	4.94	6.48
Basic ed	0.16	0.17
Some HS	0.12	0.13
HS grad	0.32	0.34
Some college	0.038	0.039
College grad	0.27	0.25
MA-PhD	0.094	0.066
Managers/professionals	0.32	0.31
Clerks/skilled workers	0.45	0.49
Other occupations	0.23	0.20
Job earnings	406.5	396.7
Part time	0.16	0.15
Peer group size	6.02	3.13
Absence days	24.4	21.8
Any absence	0.42	0.39
Observations	349,147	112,143

Note: Table shows summary statistics of treated (transferred) workers and their work peers in the main estimation sample.

The figure is constructed using data on both transferred workers and their matched peers, excluding peer groups of less than 5. The plot indicates a strong positive correlation between own and group absence. It should of course be emphasized that the pattern showed in Fig. 2 could reflect general calendar time effects, correlated characteristics of co-workers in the same firms, etc. and should not be given a causal interpretation.

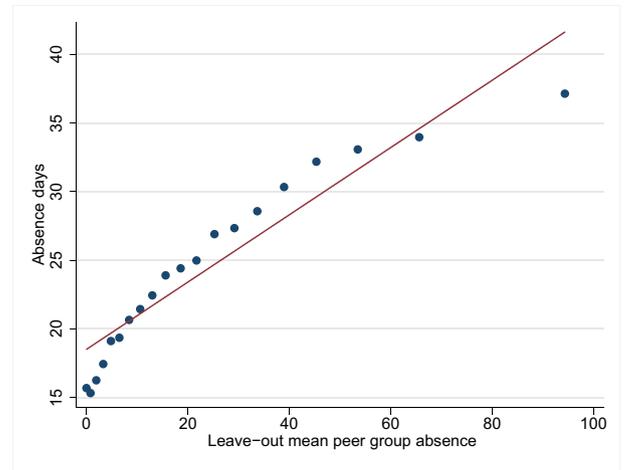
4.2. Evaluating random assignment

One central identifying assumption in our empirical strategy is that the change in doctor practice - as measured by  $\delta_i = \alpha_{p'} - \alpha_p$  - should be random. As a consequence, it should be uncorrelated with the focal worker's underlying absence propensity. We can get an indication of whether this holds by testing whether the instrument is correlated with observable worker characteristics that are linked to sickness absence (similar to the approach used in Dahl et al., 2014a).

The first exercise links the change in GP fixed effects to individual observable characteristics that may predict absence rates. To test this, we construct variables measuring the change in the focal worker's absence days and the associated doctor instrument using the year before and after the doctor transfer. These variables together with individual absence the year prior to GP change are then regressed on a set of observable characteristics - age, gender, education and job tenure - measured the year prior to GP change. Models also include municipality and year fixed effects. Results are shown in Table 2.

Column (1) shows estimated effects of observable characteristics on focal worker absence level, while column (2) shows estimated effects on the change in focal worker sickness absence. All the included covariates are highly significant in explaining variation in absence levels. Moreover, it is clear that several of the observable individual characteristics in our dataset are also significant in predicting individual absence level as well as absence changes. Being female and having higher earnings predict a larger increase in sickness absence, while more education and longer job tenure are associated with a more negative change in absence. F-test indicates that the covariates in Table 2 are jointly significant in explaining absence levels and changes.

Meanwhile, the estimated effects in column (3) are for the most part not statistically significant from zero, indicating that these covariates are not significant in predicting the change in doctor fixed



**Fig. 2.** Own and peer group absence.

Note: Figure shows binned scatter plot of annual absence days plotted against average peer group absence (constructed as leave-out mean).

effects. The one exception to this is age, which is marginally significant at the 10% level. In the context of multiple hypothesis testing, this is not unexpected. Meanwhile, F-tests fail to reject the null hypotheses that the coefficients are jointly equal to zero.

Column (4) shows the corresponding balancing test estimated on the focal worker's peers. The results from this test closely mirrors that of the focal workers. Note that age is once again significant - this follows more or less mechanically from the result for focal workers, as peers by definition are required to be no more than two years older or younger than the focal worker. Moreover, the covariates are not jointly significant, as measured by the F-test.

**Table 2**  
Testing for randomness of doctor practice change.

	(1)	(2)	(3)	(4)
	Abs days	Δ Abs	Δα	Δα, peers
Female	7.602** (0.929)	2.794** (1.151)	-0.123 (0.116)	-0.0956 (0.0741)
Earnings	-0.0265*** (0.00245)	0.00916*** (0.00297)	0.000146 (0.000260)	-0.0000643 (0.000211)
Age	0.218** (0.0381)	-0.0666 (0.0598)	0.00731* (0.00399)	0.0121** (0.00531)
HS grad	-2.436*** (0.849)	-2.489* (1.347)	0.0494 (0.0758)	0.0579 (0.0628)
Some college	-3.723** (1.623)	-2.112 (2.416)	-0.164 (0.170)	0.0491 (0.122)
Bachelor's degree	-4.532*** (1.208)	-2.712 (1.666)	0.000486 (0.110)	-0.0391 (0.104)
MA/PhD	-8.049*** (1.330)	-4.750** (2.087)	-0.0695 (0.201)	0.0785 (0.160)
Union	5.330*** (0.761)	1.007 (1.046)	-0.0405 (0.0890)	0.0372 (0.0827)
Job tenure	-0.111* (0.0619)	-0.165** (0.0827)	0.00240 (0.00551)	0.00331 (0.00492)
Managers/professionals	-1.395 (1.056)	-1.674 (1.447)	-0.0196 (0.119)	-0.0207 (0.139)
Clerks/skilled workers	-0.101 (1.100)	1.184 (1.537)	-0.0150 (0.106)	-0.0554 (0.132)
Est size < 10	-4.346*** (1.126)	2.216 (1.720)	-0.144 (0.0991)	-0.153 (0.122)
Est size > 500	1.553 (1.918)	1.214 (2.646)	-0.109 (0.212)	-0.151 (0.215)
Establishment size	-0.000969 (0.00107)	-0.000713 (0.00139)	-0.0000400 (0.000109)	-0.0000489 (0.000119)
Observations	21,462	20,514	21,462	66,249
F joint sign.	45.00	2.026	0.954	1.261

Standard errors in parentheses. Models include calendar time dummies. \*p < 0.10, \*\*p < 0.05, \*\*\*p < 0.01.

In a second set of tests, we combine the individual characteristics reported in Table 2 into a linear measure of predicted sickness absence propensity. This variable is a linear prediction of sickness absence using coefficients from a regression of sickness absence on plausibly exogenous individual characteristics: age, gender, education, union membership and job tenure. This measure is used to divide the sample into subsamples by deciles of the predicted absence distribution. We then regress focal worker absence (the variable being instrumented) and the change in GP fixed effects (the instrument) on indicator variables for deciles 2–10 of the resulting distribution, controlling for year and municipality. Fig. 3 plots estimated coefficients together with 95% confidence intervals. Higher absence propensity significantly predicts higher absence rates – this is not surprising. Meanwhile, the panel on the right indicates that there is no such relationship between absence propensity and the instrument. This lends further support to our claim of random assignment.

As discussed earlier, random assignment could be threatened if high propensity workers sort themselves to lenient GPs prior to the GP switch. This could potentially lead to the change in estimated GP effects being negatively correlated with underlying absence propensity, as these high propensity individuals are mechanically more likely to be transferred to a stricter GP. The balancing test presented in Table 2 finds that absence propensity is not significant in explaining  $\delta_i$ , indicating that this is not necessarily a problem in our data. To address this in more detail, we divide the samples of focal and peer workers into two subgroups, based on whether they moved to a slacker ( $\delta_i > 0$ ) or stricter GP. If there is a bias caused by self selection, we would expect those who are transferred to stricter GPs to have systematically higher predicted absence rates. In the appendix, we show that the absence propensity distribution is largely overlapping for the two subsamples. Moreover, the slacker/stricter subsamples are also similar in terms of observable characteristics (see Appendix Table A6). Overall, the models support our assertion that the change in doctor practice style is uncorrelated with other drivers of individual sickness absence.

## 5. Findings

### 5.1. Event-study models

Eq. (4) is estimated separately for focal workers and for their colleagues. Estimated  $\gamma^k$  with 95% confidence intervals are plotted in Fig. 4. The estimated effect for the year prior to the transfer year ( $t = -1$ ) is normalized to zero.

The left panel shows estimates for the focal workers. The graph shows a jump in the estimated  $\gamma$ s at the time of transfer. Before transfer years ( $t < 0$ ), the estimated parameters are close to zero. In year 0, the estimated parameter increases to around 0.3. (In the year of transfer, we do not know exactly which doctor handled the absence certificates.) For time  $t > 0$ , the estimated parameters flatten out with point estimates around 0.6. Recall that the instrument is constructed using only the two years immediately before and after the transfer year. The way the plot remains flat for all four years before and after then supports the interpretation of the constructed instrument as capturing some persistent measure of physician behavior rather than spurious trends in certification practices. Overall, the left panel indicates that GPs play a significant role in determining focal workers' absence patterns.

The panel on the right shows estimates for the colleagues. Estimated  $\gamma$  are flat and close to zero before  $t = 0$ . The estimated  $\gamma$  in the year of transfer is around 0.1. For  $t > 0$ , the estimated effects stabilize around 0.2. The estimates are (for the most part) statistically significant. As these are not affected directly by the transfer, we would expect there to be no increase in the estimated  $\gamma$  around the time of doctor change in the absence of peer effects. Moreover, it is important to note here that the jump is discontinuous at the time of focal worker GP transfer – the estimates do not seem to follow a trend. The spillover effects thus do not appear to be the result of spurious correlations or correlated trends. To summarize, the event-study approach supports the validity of our research design, and gives a first indication that there are significant spillover effects.

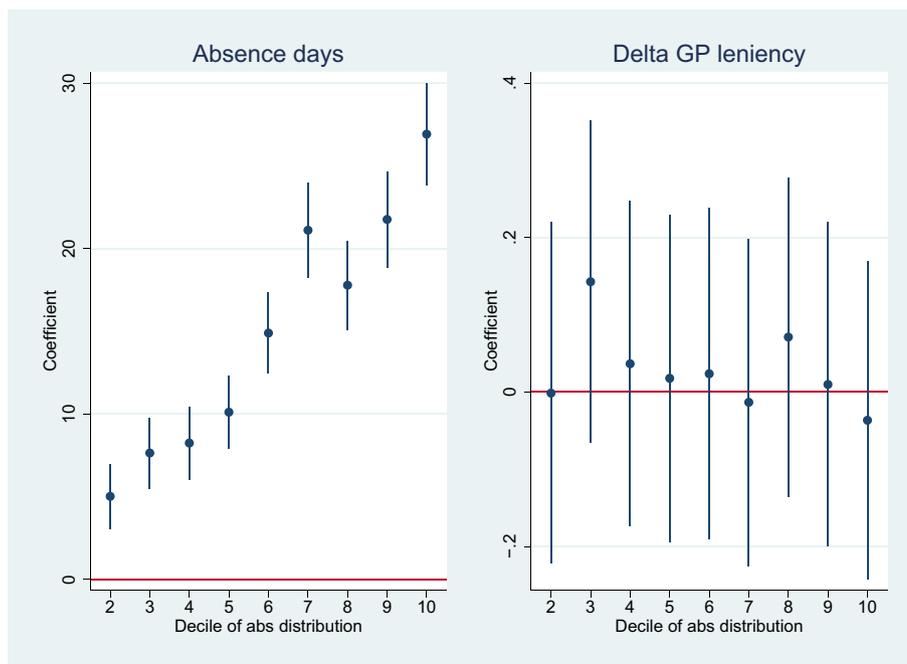
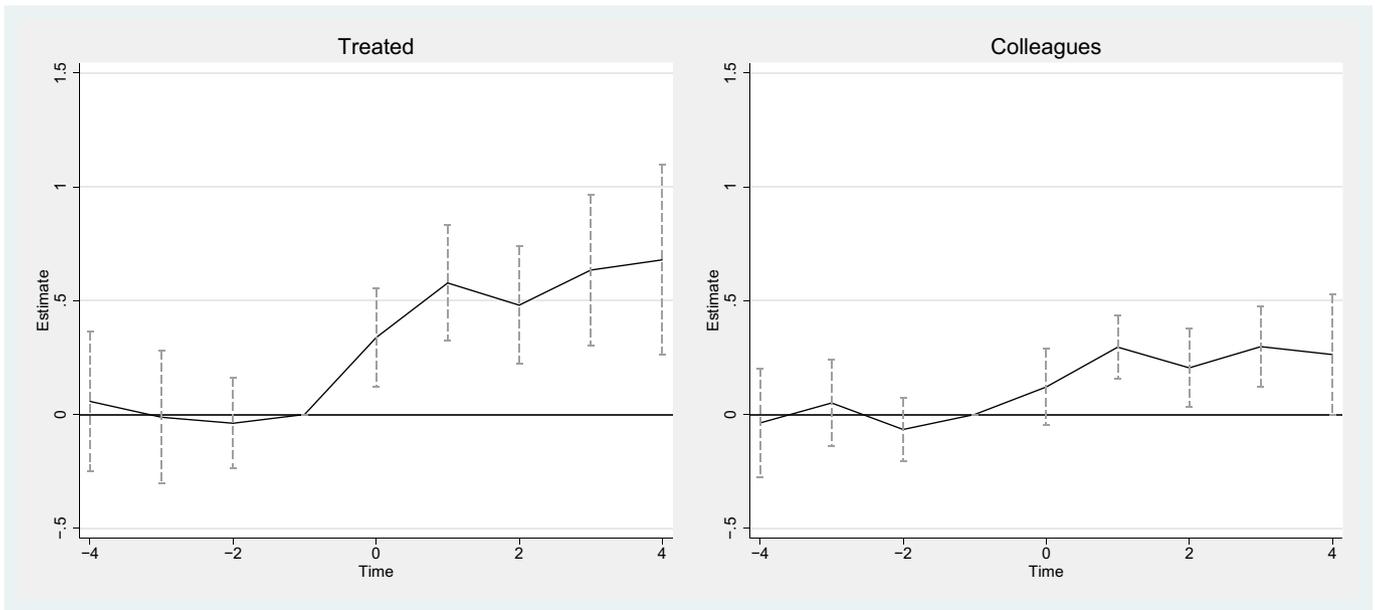


Fig. 3. Evaluating random assignment - predicted absence rates.



**Fig. 4.** Event-study plots.  
 Note: Figure plots estimated  $\gamma$  from Eq. (4). The vertical bars show with 95% confidence intervals. Effect in year  $t - 1$  normalized to zero.

5.2. IV models

The estimated doctor effects are then used to estimate panel models of peer and focal worker absence. Table 3 summarizes OLS and IV estimates of Eq. (2). First, column (1) presents the OLS estimate of the baseline model with peer group fixed effects. The estimated effect, though positive and highly statistically significant, is actually close to zero. The inclusion of fixed effects means that the estimate is identified only off year-to-year changes in focal worker absence. Dropping the peer group fixed effects from the model increases the estimated effect significantly.

Column (3) shows the estimated first stage: a one unit increase in GP leniency increases absence of the focal workers by 0.652 day. Column (4) shows the corresponding reduced form estimates of GP leniency on coworker’s absence rates. Consistent with the findings in the event-study models, we find that changes in the focal worker’s GP practice style significantly shift absence among coworkers. A one unit increase in GP leniency increases coworker absence rates by 0.265 day. The effect is highly statistically significant, indicating that GP gatekeeping has spillovers to workplace peers.

Column (5) presents the IV estimates - effects of peer absence days on coworker absence. These estimates hinge on a stronger set of assumptions than the reduced form estimates (and event study models). In particular, the IV model requires the exclusion restriction to hold.

**Table 3**  
 Model results.

	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	First	RF	IV
Abs treated	0.00670*** (0.00246)	0.0295*** (0.00244)			0.406*** (0.115)
GP leniency			0.652*** (0.118)	0.265*** (0.0531)	
Observations	349,044	349,075	349,044	349,044	349,044
F-stat					30.34
Fixed effects	Peer group	None	Peer group	Peer group	Peer group
Controls	Full	Full	Full	Full	Full

Standard errors in parentheses \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .  
 Note: Table shows OLS and IV estimates of Eq. (2). Standard errors corrected for clustering at the GP level.

In our context, this assumption is nontrivial. GPs typically make multiple decisions about patients’ care. As a consequence, the exclusion restriction may be violated if there are spillovers in health between coworkers.

To illustrate, consider a new GP who is more successful than their predecessor at improving patients’ underlying health conditions, e.g. through educating patients about health behaviors or knowledge about innovative treatments. This new GP may appear in our sample with a low estimated leniency  $\hat{\alpha}$  as their patients, now in better health, achieve lower absence rates. If there are positive spillovers in health, coworkers may achieve lower absence rates in turn not due to any spillovers in absence rates per se, but rather due to spillovers in underlying health. The IV estimate then, would be biased upward.

More generally, any violation of the exclusion restriction risks biasing the IV estimates. The available data does not allow us to test the exclusion restriction directly as we have few measures of patient health or other aspects of GP practice style. In the robustness section of the paper, we present testable implications of the exclusion restriction, which produce favorable results. Still, the biases that may occur should the exclusion restriction fail to hold is worth keeping in mind.

With this in mind, the IV estimate reported in column (5) indicates that one day’s extra absence of the treated (transferred) workers increases colleague absence by 0.406 day.<sup>19</sup> As indicated by the first stage the instrument is strong ( $F = 30$ ).

As discussed in Section 3, under the assumptions of our model, the IV estimate reported in Table 3 is the local average treatment effect on the subpopulation of compliers - those individuals whose focal worker colleague is driven to change their sickness absence as a result of the GP transfer.

These compliers cannot be individually identified in the data. In the case of a discrete treatment, we can calculate the number of compliers as well as their observable characteristics (Imbens and Rubin, 1997). In our case, where treatment is continuous (absence rate), we first discretize the treatment and outcome variables. Specifically,

<sup>19</sup> This estimate corresponds to a social multiplier of 1.59, which is in the range of comparable estimates found in the literature (Lindbeck et al., 2016; Dale-Olsen et al., 2015).

we estimate a version of Eq. (2) where the outcome and treatment variables are binary indicators equal to one if the peer and/or focal worker have any sickness absence that year, and 0 otherwise. (See Appendix for details.) The estimated effects of this exercise are consistent with the IV estimate reported in Table 3: focal worker absence increases the probability that a peer worker is absent by 40.2%. This model then forms the basis for estimating the share of compliers and their characteristics.

We estimate a complier share of 9.6%. Appendix Table A8 shows estimated complier characteristics. Compliers are similar to non-compliers in terms of gender and education. Meanwhile, compliers are more likely to be older and to work part time.

The IV estimate is very large compared to the OLS estimates.<sup>20</sup> From the literature, there are two reasons why the OLS and IV estimates may differ. First, there may be heterogeneity in effects. Second, OLS estimates may exhibit selection bias.

The selection bias will be positive if there are correlated shocks that are not accounted for in our control variables. The model fully controls for unobserved drivers of sickness absence that are constant over time through the inclusion of peer group fixed effects, as well as observable time-varying characteristics of the workers and aggregate time effects. Still, unobserved factors such as contagious infections spreading at work may induce a positive correlation in peer and focal worker absence patterns.

Heterogeneous treatment effects can also lead to IV and OLS estimates differing. The IV model estimates the local effect for complier peers. These are the individuals whose focal worker colleague changed their absence behavior as a result of their GP transfer. Given the institutional context, complier focal workers are more likely to be in a gray area of health where they are neither perfectly healthy, nor clearly too unable to work. For peer workers, changes in absence may be more salient for this group: seeing a colleague who is not entirely healthy show up at work may act as a signal of updated workplace absence norms. On the other hand, seeing a coworker who is clearly healthy be present at work, or observing that a colleague is absent from work after a serious illness or injury may not have the same impact on individual absence decisions.

### 5.3. Robustness exercises

Table 4 presents results from estimated model variations. The estimates reported in Table 3 come from models that control for age, part-time status, relative time since doctor change as well as calendar time effects. If the instrumental variable approach is valid, including controls for individual characteristics should not affect the estimated effects. Column 1 in Table 4 shows estimates from a model specification which does not include control variables (except for calendar time). This specification finds an IV estimate of 0.433, which is slightly larger, though not statistically significant from the baseline estimate of 0.406.

As discussed in Section 3, the models include a peer group fixed effect. The model is identified from within-peer group changes in the instrument; this could reflect changes in the makeup of the peer group before and after the physician change. In an alternative specification, we include fixed effects for each individual within the peer group. In this specification, peer effects are identified only from persons who worked with the treated individual both before and after the doctor change. This model is presented in column (2) of Table 4. The IV estimate changes very little (0.396 vs 0.406), indicating that composition effects are not driving our findings.

The baseline sample excludes very large peer groups, defined as age-occupation groups with 20 or more individuals. Column (3)

**Table 4**  
Model results.

	(1)	(2)	(3)
	No controls	Person FE	Keep 20+
<i>Reduced form</i>			
GP leniency	0.283*** (0.0530)	0.250*** (0.0535)	0.205*** (0.0500)
<i>IV estimates</i>			
Abs treated	0.433*** (0.116)	0.396*** (0.112)	0.355*** (0.120)
<i>First stage</i>			
GP leniency	0.652*** (0.118)	0.632*** (0.118)	0.579*** (0.130)
Observations	349,044	314,995	416,497
Kleibergen-Paap F-stat	30.30	28.51	20.01

Standard errors in parentheses \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Table shows IV estimates of Eq. (2). Column (1) shows the model estimated with no individual or firm level covariates, column (2) shows the model estimated with individual fixed effects. Column (3) shows the model estimated on a sample where large peer groups are retained. Standard errors corrected for clustering at GP pair level.

presents results when the model is estimated on a sample which includes these larger peer groups. The point estimates, while still highly significant, are reduced somewhat. Expanding the sample in this way yields a reduced form estimate of spillover effects of 0.205 and a corresponding IV estimate of 0.355, both significantly smaller than our baseline estimates. This is consistent with social interaction effects being less prominent in these larger peer groups, though it is worth noting that the instrument is also weaker in this sample.

As discussed above, IV estimation requires the exclusion restriction to hold: GP leniency should affect coworker absence rates only through changes in focal worker absence rates. Exclusion may be threatened if the estimated GP leniency indicators capture other aspects of physician practice style, such as skill or experience, which may affect focal worker health. In particular, exclusion may be violated in this case if there are resulting spillovers in coworker health.

To examine first the case of physical health spillovers, the next set of models allows for the spillover effect to vary by diagnosis category. Colleagues' absence days are grouped in four categories: conditions that affect the respiratory system, musculoskeletal conditions, psychological conditions and others. If the results are driven by physical contagion patterns, we would expect effects to be driven by respiratory conditions - that can plausibly spread between colleagues - and not by musculoskeletal or psychological conditions.<sup>21</sup>

Table 5 shows estimated spillover effects by diagnosis group. In these models, higher absence rates of the focal worker are found to have a negative effect on peer absence due to respiratory conditions. That is, the sign is reversed for this category. This finding is consistent with increasing spread of infections if sick workers are induced by the instrument to return to work early, while they are still contagious. In other words, for respiratory conditions, stricter gatekeeping on the part of the focal worker's physician may have negative spillover effects on colleagues who are exposed to cold and flu. In fact, limiting the spread of communicable diseases among colleagues in the workplace could be a motivation for paid sick leave, in order to reduce the overall prevalence of disease in the firm (Skåtun, 2003; Kumar et al., 2013; Marathe et al., 2014).

Meanwhile, point estimates of spillover effects for the other three diagnosis categories are all positive. For absence due to psychological conditions, the IV estimate is not statistically significant, while the reduced form estimate is significant only at the 10% level. For

<sup>20</sup> A similarly small OLS estimate relative to the IV estimate is found in Dahl et al. (2014a).

<sup>21</sup> Many common respiratory conditions are transmitted by airborne pathogens, making it easier for this group of infections to spread among colleagues at work.

**Table 5**  
Effects by diagnosis category.

	(1)	(2)	(3)	(4)	(5) First stage
	Respiratory	Psychological	Musculoskeletal	Other	First stage
<i>Reduced form</i>					
GP leniency	-0.0197** (0.00979)	0.0404* (0.0242)	0.158*** (0.0351)	0.0863*** (0.0302)	
<i>IV estimates</i>					
Abs treated	-0.0302* (0.0157)	0.0620 (0.0396)	0.242*** (0.0726)	0.132** (0.0516)	
<i>First stage</i>					
GP leniency					0.652*** (0.118)
Observations	349,044	349,044	349,044	349,044	349,044
Kleibergen-Paap F-stat	30.34	30.34	30.34	30.34	
$\bar{y}$	1.518	4.069	10.20	8.582	

Standard errors in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Note: Table show IV estimates of Eq. (2) where the dependent variable is days of sickness absence in various diagnosis categories, based on ICPC classifications. Standard errors corrected for clustering at the GP level.

the other two groups, effects are highly significant. Quantitatively, effects appear to be roughly proportional to prevalence of each absence category. The fact that the spillover effect is not driven by communicable diseases would appear to lend further support a behavioral interpretation of spillover effects.

To further investigate the exclusion restriction, we estimate augmented models where additional GP observable characteristics are included as controls. We find indications that these characteristics are significant in explaining focal worker absence. If the exclusion restriction holds, including these additional controls should not affect the IV estimate.<sup>22</sup> Results are shown in Appendix Table A9. Meanwhile, the IV estimate does not change significantly.

Monotonicity requires that the causal channel works only in one single direction. A testable consequence of this assumption is that the first stage should be non-negative for all subgroups. Tables A10 and A11 in the Appendix present results for the estimated first stage estimated for each quartile of the absence propensity distribution as well as for different diagnosis groups. The estimated point estimates are all positive, indicating that monotonicity holds.

In our baseline specification, colleagues are dropped from the sample once they leave the focal worker's peer group; moreover, when a worker focal worker leaves the sample, they and their peers are removed from the sample. This ensures that peers and focal workers are more likely to have social interactions, however, this choice does significantly reduce sample size, as job changes or even changes in occupational status lead observations to be cut from the estimation sample. To see how this choice impacts our results, we construct an alternative sample where peer groups are locked at  $t^* - 1$ , retaining both peer and focal workers for all years with non-missing observations, ignoring changes in employer or occupational status. Results are presented in Appendix Table A2. As expected, the sample size is significantly larger - 711,022 vs our baseline 349,044. Point estimates are significantly smaller - effects will likely be diluted as the sample now includes "peers" who may no longer actually be working together with the focal worker, or even in the labor force. Estimated effects remain highly statistically significant.

By necessity, we make a number of admittedly arbitrary decisions when constructing the auxiliary sample defining GP transfers. In particular, we require that at least 85% of the incumbent GP's patients transfer to the replacement GP. To evaluate the robustness of our findings to these choices, we estimate a set of models varying the threshold required to define a transfer from 50 to 95%. Results from this exercise are shown in Appendix Table A3. Qualitatively, results are robust to these variations - estimates are positive and significant for a range of thresholds.<sup>23</sup> Setting the threshold lower reduces the point estimate compared to our baseline specification. For the lowest level we consider, 50%, the resulting IV estimate of spillovers is 0.293, a 28% reduction from our baseline estimate of 0.406. The relative imprecision of these estimates means that this difference is not statistically significant. The reduced form spillover estimates meanwhile are significantly smaller when the threshold is lower.

Finally, we want to make sure that the estimated effects represent peer effects and not some correlated occupation-specific trends in absence patterns. To examine this, we estimate a set of placebo models where we would not expect to find any spillover effects if the assumptions of our model hold.

In the first specification, the empirical models are estimated using samples of focal workers matched with a set of pseudo-colleagues in other firms. If the estimated effects indeed represent peer effects in absenteeism, we would expect effects to occur only for persons who work together. To construct a sample of pseudo-colleagues, the focal workers are randomly matched to colleagues in a different firm in the same industry, located in the same county, employed in the same occupation. The IV model is then estimated on this sample of pseudo-peers. If the estimates discussed above represent correlated trends around doctor changes, we would expect there to be significant effects of peer absence in this placebo regression. However, if there truly are spillover/peer effects, we would expect zero effects on this group, as there is (presumably) no social interaction.

Estimated effects of this exercise are shown in columns (1)–(3) of Table 6. In these models, the first stage is similar to what we found in our preferred specification. However, the model finds no effects on the pseudo-peers. The absence of spillover effects in groups where

<sup>22</sup> Specifically, we include physician age (5 categories), gender and a dummy variable indicating whether the GP and patient are of the same gender. While older GPs are more experienced, younger GPs who are recent graduates may be more up-to-date in their medical knowledge. Gender may have an impact on treatment if patients are more receptive to a GP of their own gender.

<sup>23</sup> Setting the threshold as high as 90%, the instrument is no longer strong enough to meaningfully run IV, likely due to the high number of observations excluded/low power.

**Table 6**  
Placebo model results.

	Pseudo-colleagues			Pseudo-shifts
	(1)	(2)	(3)	(4)
	RF	First	IV	RF
GP leniency	-0.0200 (0.0742)	0.494*** (0.126)		0.0905 (0.0765)
Abs treated			-0.0404 (0.151)	
Observations	150,802	150,802	150,802	242,654
Kleibergen-Paap F-stat			15.30	

Standard errors in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Note: Table show results of placebo models: Columns (1)–(3) show selected coefficients from models estimated on a sample of focal workers linked to pseudo-peers (in other firms). Column (4) shows reduced form estimates of the models estimated on a sample of pseudo-shifts, where the instrument is constructed using individuals who did not change doctors. Standard errors corrected for clustering at the GP level.

we suspect correlated shocks (e.g. local business cycles) but no social interaction supports the validity of the GP instrument.

The second placebo further assesses the identification strategy's robustness to local shifts in sickness absence. Local unobserved factors such as an above-average local seasonal influenza outbreak could raise absence rates across the local community (relative to the average that year). This would increase both the observed absence rates of workers in local firms and patients of local GPs. Such a shift could lead to a positive correlation between estimated GP effects and peer group absence rates, leading to a spurious correlation, threatening our identification strategy. If this is the case, we should see a similar significant spillover effect even if there never was an actual GP transfer.

In the second placebo model, we address this by selecting one non-treated colleague from each peer group.<sup>24</sup> Then, the instrument is constructed as if the pseudo-focal worker changed doctors at time  $t_j^*$ , even though no such change took place (a "pseudo-shift"). In a first step, a placebo instrument is estimated using the auxiliary model outlined in Section 3. That is, we re-estimate the physician fixed effects as if the peer worker changed GP at time  $t_j^*$ , while in fact, the GP did not change. As there is no true physician shift, the difference between the pair of estimated physician effects  $\alpha_p - \alpha_{p'}$  now captures differences in average absence certification for the same doctor in two different time periods.

Results from this exercise is shown in column (4).<sup>25</sup> The model finds no spillover effects of this pseudo-shift on peer workers. Taken together, these two placebo models find no evidence that the exclusion restriction is violated, further supporting our findings.

## 6. Spillovers - mechanisms and extensions

### 6.1. Effects by peer group characteristics

By instrumenting focal worker absence rates, we ensure that the estimated effects on coworker sickness absence capture causal spillover effects rather than the results of endogenous sorting or correlated shocks. In this section, we estimate a set of models making the case that the estimated effects in fact reflect social spillovers in absence behavior. Next, extended models are estimated to shed further light on the social transmission mechanisms.

In the case of social spillover effects, we would expect estimated effects on peer absence to be stronger among colleagues who interact

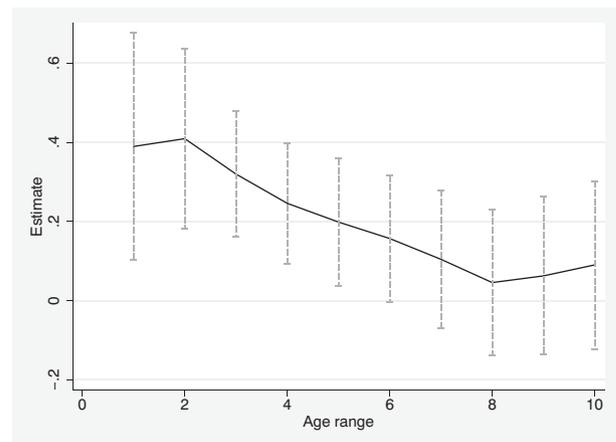
more at work. While the degree of interaction between colleagues is not directly observable, we can observe whether effects are stronger when peer workers are more similar to each other. In the main estimation sample, peer groups are defined using a two-year age bracket around the focal worker. To see how this choice of age bracket affects our estimates, alternative estimation samples are constructed using different cutoffs for the maximum allowable age range of peer workers, from  $\pm 1$  to  $\pm 10$ . When interpreting these estimates, we should keep in mind that changing the peer group definitions is likely to change other peer group characteristics, for example, average peer group size increases with age bracket. Smaller peer groups may have more social interactions, regardless of individuals' proximity in age.

Model (2) is then estimated on each sample. Resulting IV estimates are plotted in Fig. 5. Estimated effects are the largest when colleagues are required to be relatively close in age to the focal worker. When the estimation samples allow for larger age gaps, estimated spillover effects are reduced, becoming not statistically significant at age range  $\pm 7$ .

As a related exercise, we construct samples using alternative definitions of occupational groups. While peer groups in the preferred estimation sample are defined using occupation classifications at the 3-digit level - in alternative samples, broader groups are constructed using 1- and 2-digit levels. In these models (not shown) estimates remain significant and positive. Point estimates are reduced somewhat, though the impact is limited (with a 2-year allowable age gap, point estimate is reduced from 0.41 to 0.31 when moving from a 3-digit to a 1-digit occupation category).

One reason why employees may want to adhere to a local absence norms is that they fear "excessive" absence may have adverse career consequences. There is evidence that job protection may increase absenteeism (Ichino and Riphahn, 2005). Moreover, absenteeism tends to be positively correlated with labor market tightness, consistent with a disciplining effect of unemployment on absence rates (Askildsen et al., 2005). From this literature, we could expect workers to be less willing to deviate from local absence norms when faced with a perceived threat of job loss - predicting stronger spillover effects when job security is low. However, if employees are more eager to avoid unnecessary absenteeism when job security is low, we might see the reverse - spillover effects might be smaller when the risk of layoff is high.

To examine this, the next set of models examines how the estimated spillover effects vary with individuals' risk of job loss. Proxies for job security are constructed using a setup similar to Dahl et al. (2014b). Specifically, we identify observations with low unionization



**Fig. 5.** Estimates by peer worker age bracket.

Note: Figure plots estimated  $\gamma$  from Eq. (2), estimated on estimation samples with different age brackets. Each plotted estimate represents a separate regression.

<sup>24</sup> We have chosen the individual closest to the focal worker in age.

<sup>25</sup> The first stage and IV estimates are not reported - while the first stage is positive, the instrument is not strong enough to run IV. Consequently, only the reduced form estimate is reported.

**Table 7**  
Job security.

	(1)	(2)	(3)	(4)
	High unempl	Low union	Low unempl	High union
<i>Reduced form</i>				
GP leniency	0.269*** (0.0818)	0.376*** (0.0976)	0.205** (0.0931)	0.232*** (0.0731)
<i>IV estimates</i>				
Abs treated	0.443** (0.190)	0.751** (0.342)	0.336* (0.194)	0.311*** (0.119)
<i>First stage</i>				
GP leniency	0.607*** (0.169)	0.500*** (0.185)	0.610*** (0.181)	0.746*** (0.162)
Observations	163,494	113,829	181,889	234,232
Kleibergen-Paap F-stat	12.84	7.273	11.41	21.32

Standard errors in parentheses \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Note: Table shows IV estimates of Eq. (2). Columns (1) and (2) show estimates from “low job security” samples: Unemployment > 2.5%, unionization rates < 70%. Columns (3) and (4) show estimates from corresponding “high job security” samples. Standard errors corrected for clustering at the GP level.

rates (less than 70%) and high local unemployment rate (registered unemployment rates less than 2.5%).<sup>26</sup>

Results are shown in Table 7. Estimated effects are slightly smaller for low job security subsamples. In particular, effects are larger for firms with lower unionization rates, though the weakness of the instrument in this subsample makes interpretation difficult. Meanwhile, estimates remain significant also for the high job security subsamples. Overall, the estimates in Table 7 indicate that while the threat of job loss may be a factor explaining peer effects in absenteeism among colleagues, peer pressure also appears to be a factor in controlling absence rates for workers who are secure in their jobs.

Next, additional models have been estimated to shed some light on two alternative mechanisms. First, there may be social interaction effects in the physician switch itself. Co-workers may respond to the focal worker's physician transfer by changing their own physician affiliation, for instance if the focal worker transmits some information about the GP system. In particular, if the focal worker is transferred to a doctor who is more willing to certify absences, peer workers may be inspired to seek out a doctor who is more willing to comply with requests for absence certificates. This kind of asymmetric effect could be problematic for the interpretation of our IV estimates, as the exclusion restriction might be violated. In extended models (not shown), we find that while colleagues of the transferred workers are significantly more likely to voluntarily switch doctors after the focal worker is transferred. However, this increase happens regardless of whether the focal worker is transferred to a stricter or slacker doctor.<sup>27</sup>

In addition, there may be spillovers in terms of physical or mental effort as well as social interaction effects. In demanding jobs, colleagues may experience increased strain when the focal worker's absence increases, which in turn could lead to higher absence rates as the affected colleagues become more exposed to excess psychological or physical stress. To shed more light on this possible mechanism, the sample is merged to an occupation-level dataset with information on various aspects of working conditions, and instrumental

variable models estimated separately for more and less physically demanding jobs.<sup>28</sup>

Overall, there does not seem to be a systematic pattern. While high workload occupations exhibit larger peer effects, low support occupations exhibit relatively smaller peer effects. Workers in more physically demanding occupations do not appear to be more vulnerable to absence spillovers than those in less physically demanding jobs. Moreover, the estimated effects are imprecisely estimated, so we cannot conclude that they are significantly different from one another.

Finally, we have estimated a set of models to investigate the question of symmetry, that is, whether peer effects in absenteeism differ depending on whether the GP transfer induces an increase or a reduction in focal worker absence. From economic theory, it is unclear whether we should expect to see symmetrical results. In Kandel and Lazear (1992)'s models of peer pressure, workers suffer a penalty when deviating from local effort norms. However, whether the penalty is incurred symmetrically for deviations in either direction of the norm depends on the parameters of the model, which again are likely to depend on context.

Table A16 in the appendix shows selected coefficients of the model estimated separately on subsamples where the focal worker was transferred to a slacker/stricter GP. The models estimate stronger peer effects when focal workers are transferred to a GP with a lower certification propensity (a “stricter” doctor), however, the effects are not statistically significantly different from each other.<sup>29</sup>

## 6.2. Policy simulation

Our findings indicate that policies that affect sickness absence will have indirect impacts through spillover effects among colleagues. This is likely to be the case for a wide range of policies, such as reducing or increasing the generosity of sick pay benefits. The quantitative implications of our findings are clearest in the case of policies that affect gatekeeping. To illustrate this, we construct a policy simulation following the approach of Dahl et al. (2014b). Specifically, we consider a policy that successfully reduces doctor leniency by 1/5 of a standard deviation.

To estimate these effects, we first regress sickness absence on the physician leniency indicators, controlling for year, peer group fixed effects and time-varying observables:

$$y_{it} = \theta_j + \theta_t + x_{it}\beta + \hat{\alpha}_{p(j,t)}\Gamma + \varepsilon_{it}$$

The model is estimated separately for the focal workers and peers, and correspond to the first stage and reduced form models in Table 3. The estimated  $\hat{\Gamma}$  are then multiplied by one fifth of the standard deviation of the change in physician practice indicators in the sample. This gives the predicted reduction in absence days for each group. Dividing this by the average absence days for each group gives the predicted reduction in absence rates.

Table 8 presents the results from this exercise. As expected, effects are larger for the directly affected workers. On average, a 0.2 standard deviation reduction in physician leniency translates to 0.318 fewer absence days per focal worker, or a 1.5% reduction relative to the sample mean. The peer workers have 0.146 fewer absence days, or a 0.6% reduction relative to the sample mean.

The last row of Table 8 shows the weighted average of these effects: on average, absence rates in the sample are reduced by 0.8%.

<sup>26</sup> Municipal unemployment figures from Statistics Norway. Unemployment was low throughout the period covered in this paper, the sample average local unemployment rate is 2.6%.

<sup>27</sup> Moreover, results are robust to excluding peer workers who changed GPs over the period.

<sup>28</sup> We use data from Statistics Norway's survey on working conditions and working environment carried out in 2013.

<sup>29</sup> Moreover, weak instruments in the subsamples complicate the interpretation of these differences.

**Table 8**  
Policy simulation.

Group	(1)	(2)	(3)
	Days	Rate	Share
Treated ( <i>N</i> = 112,141)	0.318*** (0.049)	0.015	43.9%
Peers ( <i>N</i> = 349,075)	0.146*** (0.029)	0.006	56.1%
All	0.188	0.008	100.0%

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Note: Table summarizes results from a within-sample policy simulation of a 0.2 standard deviation reduction in physician leniency.

Focal workers account for only 43.9% of the effects of the policy: spillover effects on peers account for 56.1%.

## 7. Conclusions

Absence rates tend to be correlated within peer groups at work. In this paper, we examined whether this correlation reflects spillover effects in absenteeism. To identify causal spillover effects, we exploited shifts in absence probabilities that occur when patients are shifted between physicians with potentially different certification behavior. Individuals who are made to transfer between doctors in this way experience significant changes in their own absence rates around the time of the doctor change. However, in the absence of spillover effects between colleagues, there would be no reason to expect the doctor transfer to affect the absence patterns of their non-treated work peers.

Using this identification strategy, we find significant spillover effects in absenteeism: There is a significant shift in absence patterns of non-treated work peers around the time when the focal workers are transferred between doctors. In IV models, we find that one additional absence day of the focal workers increases peer absence by 0.41 absence days. In a policy simulation, we find that these spillovers significantly amplify the impact of tightened gatekeeping: we estimate that spillovers will explain 56% of the total policy impact.

The public discussion on absenteeism frequently refers to social norms with respect to absence – the workplace is a key arena where such norms are likely to form and be changed. Moreover, changing gatekeeping institutions, in particular making it harder to receive sickness benefits, is a topic of contention in many countries.

The distinction of whether or not a patient is too sick to work is not always clear cut, leaving considerable discretion to the GPs. In an effort to reduce sickness absence rates, policies have been suggested that would limit doctor certification practices, i.e. by standardizing certification periods for certain diagnoses or requiring a second opinion. The presence of spillover effects in absenteeism identified in the present paper could give rise to multiplier effects in the impact of such policies. Increased gatekeeping will not only have a direct effect on the absence rates of the affected patients, but will also reduce absence rates of their peers at work.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpubecon.2018.08.015>.

## References

- Adams, J.S., 1963. Towards an understanding of inequity. *J. Abnorm. Soc. Psychol.* 67 (5), 422.
- Adams, J.S., Rosenbaum, W.B., 1962. The relationship of worker productivity to cognitive dissonance about wage inequities. *J. Appl. Psychol.* 46 (3), 161.
- Akerlof, G.A., 1982. Labor contracts as partial gift exchange. *Q. J. Econ.* 543–569.
- Akerlof, G.A., Yellen, J.L., 1990. The fair wage-effort hypothesis and unemployment. *Q. J. Econ.* 255–283.
- Angrist, J.D., 2014. The perils of peer effects. *Labour Econ.* 30, 98–108.
- Askildsen, J.E., Bratberg, E., Nilsen, Ø.A., 2005. Unemployment, labor force composition and sickness absence: a panel data study. *Health Econ.* 14 (11), 1087–1101.
- Åslund, O., Fredriksson, P., 2009. Peer effects in welfare dependence quasi-experimental evidence. *J. Hum. Resour.* 44 (3), 798–825.
- Blomqvist, Å., 1991. The doctor as double agent: information asymmetry, health insurance, and medical care. *J. Health Econ.* 10 (4), 411–432.
- Casadesus-Masanell, R., 2004. Trust in agency. *J. Econ. Manag. Strateg.* 13 (3), 375–404.
- Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., Yagan, D., 2011. How does your kindergarten classroom affect your earnings? Evidence from project star. *Q. J. Econ.* 126 (4), 1593–1660.
- Dahl, G.B., Kostol, A.R., Mogstad, M., 2014a. Family welfare cultures. *Q. J. Econ.* 1711, 1752.
- Dahl, G.B., Løken, K.V., Mogstad, M., 2014b. Peer effects in program participation. *Am. Econ. Rev.* 104 (7), 2049–2074.
- Dale-Olsen, H., Østbakken, K.M., Schone, P., 2015. Imitation, contagion, or exertion? Using a tax reform to reveal how colleagues' sick leaves influence worker behaviour. *Scand. J. Econ.* 117 (1), 57–83.
- Dusheiko, M., Gravelle, H., Jacobs, R., Smith, P., 2006. The effect of financial incentives on gatekeeping doctors: evidence from a natural experiment. *J. Health Econ.* 25 (3), 449–478.
- Fehr, E., Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. *J. Econ. Perspect.* 14 (3), 159–181.
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *Q. J. Econ.* 817–868.
- Finkelstein, A., Gentzkow, M., Williams, H., 2016. Sources of geographic variation in health care: evidence from patient migration. *Q. J. Econ.* 131 (4), 1681–1726.
- French, E., Song, J., 2014. The effect of disability insurance receipt on labor supply. *Am. Econ. J. Econ. Pol.* 6 (2), 291–337.
- Grytten, J., Sørensen, R., 2003. Practice variation and physician-specific effects. *J. Health Econ.* 22 (3), 403–418.
- Hesselius, P., Nilsson, J.P., Johansson, P., 2009. Sick of your colleagues 'absence'? *J. Eur. Econ. Assoc.* 7 (2–3), 583–594.
- Ichino, A., Riphahn, R.T., 2005. The effect of employment protection on worker effort: absenteeism during and after probation. *J. Eur. Econ. Assoc.* 3 (1), 120–143.
- Imbens, G.W., Rubin, D.B., 1997. Estimating outcome distributions for compliers in instrumental variables models. *Rev. Econ. Stud.* 64 (4), 555–574.
- Kandel, E., Lazear, E.P., 1992. Peer pressure and partnerships. *J. Polit. Econ.* 100 (4), 801–817.
- Kumar, S., Grefenstette, J.J., Galloway, D., Albert, S.M., Burke, D.S., 2013. Policies to reduce influenza in the workplace: impact assessments using an agent-based model. *Am. J. Public Health* 103 (8), 1406–1411.
- Lindbeck, A., Palme, M., Persson, M., 2016. Sickness absence and local benefit cultures. *Scand. J. Econ.* 118 (1), 49–78. <https://doi.org/10.1111/sjoe.12131>.
- Maestas, N., Mullen, K.J., Strand, A., 2013. Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. *Am. Econ. Rev.* 103 (5), 1797–1829.
- Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. *Rev. Econ. Stud.* 60 (3), 531–542.
- Marathe, A., Chen, J., Eubank, S., Liao, S., Ma, Y., 2014. Impact of paid sick leave policy: a social planner's perspective. *Am. J. Public Health* 104 (1), e1–e1.
- Markussen, S., Røed, K., 2015. Social insurance networks. *J. Hum. Resour.* 50 (4), 1081–1113.
- Markussen, S., Røed, K., 2017. The market for paid sick leave. Technical report.
- Markussen, S., Røed, K., Røgeberg, O., 2013. The changing of the guards: can family doctors contain worker absenteeism? *J. Health Econ.* 32 (6), 1230–1239.
- Markussen, S., Røed, K., Røgeberg, O.J., Gaure, S., 2011. The anatomy of absenteeism. *J. Health Econ.* 30 (2), 277–292.
- OECD, 2010. Sickness, Disability and Work: Breaking the Barriers. Organisation for Economic Co-operation and Development.
- Rege, M., Telle, K., Votruba, M., 2012. Social interaction effects in disability pension participation: evidence from plant downsizing\*. *Scand. J. Econ.* 114 (4), 1208–1239.
- Skåtun, J.D., 2003. Take some days off, why don't you?: endogenous sick leave and pay. *J. Health Econ.* 22 (3), 379–402.