The market for paid sick leave

By

Markussen, Simen and Knut Røed

July 6, 2017

The Market for Paid Sick Leave

Simen Markussen and Knut Røed*

The Ragnar Frisch Centre for Economic Research

Abstract

In many countries, general practitioners (GPs) are assigned the task of controlling the validity of their own patients' insurance claims. At the same time, they operate in a market where patients are customers free to choose their GP. Are these roles compatible? Can we trust that the gatekeeping decisions are untainted by private economic interests? Based on administrative registers from Norway with records on sick pay certification and GP-patient relationships, we present evidence to the contrary: GPs are more lenient gatekeepers the more competitive is the physician market, and a reputation for lenient gatekeeping increases the demand for their services.

JEL classification: H55, I11, I18

Keywords: absenteeism, gatekeeping, competition, role-conflicts

* This paper is part of the project "Disentangling absence patterns" financed by the Norwegian Research Council (research grant no. 227103). We wish to thank Bernt Bratsberg, Magne Mogstad, and seminar participants in Oslo and Aarhus for comments/discussions, and three anonymous referees for constructive criticism and a number of useful suggestions. Corresponding author: Simen Markussen (simen.markussen@frisch.uio.no).

1 Introduction

In most OECD countries, general practitioners (GPs) have been entrusted a role of huge fiscal importance: To protect the public purse from unwarranted social insurance expenditures and unnecessary (or cost-ineffective) medical treatments. In particular, they have been assigned the task of certifying sick pay and disability insurance expenditures; i.e., to decide whether or not (and for how long) a given health problem justifies insurance payouts; see, e.g. Bonato and Lusinyan (2007) and OECD (2010). This appears to be a logical and rational solution to a moral hazard problem. Given that work – at least for some employees – entails elements of disutility, there is a temptation to exaggerate or even "invent" health problems that can justify paid absence. The GP is probably the person who is best placed to objectively assess the true health condition of the patient, and thus decide on the need for a sick leave. And GPs tend to be highly respected and trusted citizens, to whom we presumably safely can entrust the difficult task of balancing the needs and desires of their own patients against the public costs and common interests.

Or can we? Deciding on the need for, e.g., paid sick leave obviously entails subjective judgment. Existing evidence has shown that there is considerable variation in the way GPs interpret and perform their gatekeeper role; see, e.g., Wilkin (1992), Grytten and Sørensen (2003), and Markussen et al. (2011; 2013). Patients are generally free to choose their GP, and can potentially substitute a lenient GP for a strict one. Moreover, in most countries, the GPs operate businesses in a competitive environment; hence, they may have financial incentives to attract and retain customers by providing (excess) access to publically paid services, treatments, and insurance payments. In essence, the GPs have been assigned the task of protecting the public (or private) insurer's purse against the customers who form the basis for their own livelihood. We normally think of a high degree of competition as desirable market characteristic, resulting in better services and lower prices. However, when some of the offered "products" are paid for by a third party (e.g., the taxpayer) more competition may also imply more waste. By combining the apparently incompatible roles of customer competition and gatekeeping, GPs may have been assigned a "mission impossible", in the sense that GPs who perform their gatekeeping role as intended by their principal may be forced out of the market.

The aim of the present paper is to examine empirically the practical consequences of this potential role conflict, in terms of the GPs' choices of gatekeeping standards and the workers' choices of GPs. The paper consists of three separate, but closely related, parts. First, we seek to identify each primary care physician's degree of gatekeeper leniency at each point in time. This is done month by month based on observed absence certification for workers on the patient list of the physician, after controlling for customer composition by means of worker fixed effects. Identification of the physicians' behavior derives from frequent movements of workers between family doctors as well as by their movements between employment (exposure to the risk of absence certification) and non-employment (no exposure). Second, we examine the extent to which workers choose family doctors with an eye to their reputation on gatekeeper leniency. This is done within the framework of a conditional logit model, where the choice set of available GPs is identified from the observed GP choices among other people in the local area, and the GPs' presumed leniency reputations are derived from past leniency indicators estimated on the basis of *other* employed customers only. We also study the decision to move away from a existing family doctor, and how it depends on changes in the current doctors' leniency. Finally, we examine the extent to which physicians adjust their gatekeeper leniency in response to fluctuations in the demand for their services or in the cost of losing customers. This is done within the framework of a fixed effects model where the causal effects are identified solely on the basis of changes in the local competitive environment or in the physician's remuneration structure.

We are not aware of existing empirical research on the role of physician leniency for the patients' choices of GPs. Neither is there a rich literature on the extent to which GPs take personal economic motives into account in the performance of an explicitly assigned gatekeeper role. There is a substantial literature on the impacts of economic incentives on physician behavior more generally, however, showing that payment design can have a large effect on the physicians' prescription of medical treatments; see McClelland (2011) and Chandra et al. (2012) for recent reviews. There is also more direct evidence related to GPs referral practices. For example, based on a reform in the funding scheme for certain types of elective surgery in the UK, Dusheiko et al. (2006) show that GPs tend to have a more restrictive referral practice when money *not* spent on surgery alternatively can be channeled back to their own practice. Iversen and Lurås (2000) and Iversen (2004) show that Norwegian physicians who experience shortage of customers provide more services and thus obtain higher income *per customer* than their unconstrained colleagues, whereas Iversen and Ma (2011) show that more intense local competition between physicians leads to more diagnostic radiology referrals. For referrals to specialist treatment that substitute for own services, Godager et al. (2015) note that increased competition between primary care physicians has ambiguous effects on gatekeeping incentives. While high competition and/or fewer patients than desired makes it potentially costly – in terms of lost customers – to be a strict gatekeeper, the lack of patients also makes it more profitable to treat the patient within own practice. Godager et al. (2015) provide empirical evidence that these effects largely cancel out, and that competition among GPs has insignificant (or slightly positive) effects on GPs' referrals of patients to specialists. Pike (2010), however, presents evidence from England indicating that higher competition (measured by the number of closely located rival GP practices) reduces the number of unwarranted hospital referrals.

For absence certification practices, no such ambiguities exist. Unless there is excess supply of patients, the physician's economic incentives unequivocally point toward going along with the wishes of their customers. However, there is to our knowledge little empirical evidence on the effect of physician incentives on absence certification. The only study we have found on this topic has been conducted by the Swedish Social Insurance Inspectorate (Inspektionen för socialförsäkringen, 2014) on the basis of a reform in 2007 whereby county-administrations were given the opportunity to scale down GP entry barriers and increase their inhabitants' freedom to choose their (absence-certifying) GP. As a result, the degree of competition increased considerably in some – but not all – counties, and this natural experiment is exploited to assess empirically the impacts of the GPs' economic incentives on their absence certification behavior. Based on a difference-in-difference identification strategy, the authors conclude that the number of certified absence spells increased by approximately 3.5 % as a direct result of the intensified competition between physicians.

In this paper, we present strong and robust empirical evidence that workers do take the physicians' gatekeeping reputation into account when selecting (or deselecting) a family doctor. If a physician's reputation changes such that the average worker can expect to be granted one extra day of monthly certified absence (corresponding to a movement from around the 10^{th} to the 90^{th} percentile in the estimated GP reputation-distribution), the relative probability of being chosen over each of the competing GP alternatives increases by approximately 14 %. We also present evidence indicating that many GPs take this demand curve into account by adjusting their gatekeeping practices in response to changes in their own vacancies and/or in the local competitive situation. A larger number of vacant patient slots and/or increased competition among physicians in the local residential area induce the average GP to become more lenient. Using a group of GPs on fixed-wage contracts as a point of reference, we estimate that these two mechanisms together are responsible for raising the overall level of absenteeism in Norway by approximately 4 %. This is a significant, though not a huge, impact. Yet, we argue that it could have been considerably bigger had it not been for the low level of competition between family doctors in Norway.

Although we focus exclusively on absence certification in this paper, given that this is the kind of gatekeeper decisions for which we have good administrative data, we emphasize that our findings related to the impacts of the workers' choice of GP and the GPs' choice of leniency may reflect other aspects of the GPs gatekeeping practices as well. Physicians that are lenient with respect to absence certification may also be more inclined to adhere to the wishes of their patients more generally, also including specialist referrals, drugs prescriptions, and disability insurance certification; hence the behavioral impacts identified in this paper, as well as the consequences of them, need to be interpreted in this light.

2 Institutional setting and data

In Norway, workers receive 100 % wage compensation from the first day of sickness absence and for up to one year (up to an income ceiling of approximately NOK 530,000 ($ 62,000) p.a. (2015)).[1] The first 16 days are paid for by the employer, after which the costs are covered by general (payroll) taxation. To offset moral hazard problems, a sick leave certificate issued by a physician is normally required for all spells lasting more than three days.[2] On a typical working day, 6-7% of Norwegian employees are absent from work due to sickness, and almost 90% of these absences are certified by a physician.

Norway has, since May 2001, practiced a family (panel) doctor system, whereby each citizen is assigned a single GP. In most cases, the family doctor receives a capitation fee from the social security administration (SSA) in addition to a per-treatment-pay, which is shared between the patient and the SSA. The capitation fee is currently (2015/2016) NOK 427 (approximately $ 50) per year, and it is paid out for each individual on a family doctor's customer list, regardless of whether the individual ever shows up at the GP's office or not (approximately one third of the customers do not show up during a year). Each GP determines a desired (maximum) number of list-members, with an upper limit of 2,500. The additional per-treatment pay schedule is then regulated through negotiations between the Norwegian Medical Association and the state. A standard 15-20-minute daytime consultation is currently charged at NOK 143 ($ 17). Both the

[1] To translate NOK amounts to USD, we have used the exchange rate as of November 2015, with $1=NOK 8.60.

[2] Some firms have agreed to accept self-reported sickness claims for up to eight days.

capitation fee and the per-treatment pay imply that GPs have financial incentives to keep their patients happy.

The payment system described above does not apply for all the family doctors, however. Around 4.5 % of the doctors are employed by the municipality in fixed-wage contracts, independently of the number of patients and treatments. Such fixed-wage contracts are typically offered in rural areas where it may be difficult to establish sufficiently profitable family doctor businesses, or where potential GP candidates are either risk averse or financially constrained.

Sickness absence certificates can in principle be issued by any authorized physician, but will normally be issued by the family doctor, except in emergency cases (outside the family doctor's opening hours and when the family doctor is busy) and when the patient is hospitalized or subject to specialist treatment. Norwegian workers are free to choose their family doctor insofar as the desired doctor has vacant patient slots. This can easily be done on a user-friendly web-portal, but it is not possible to change family doctor more than two times per year. When a GP has reached the desired patient ceiling, it is in principle no longer possible to choose that doctor, although there are some exemptions from this rule. Persons who do not make an active choice are assigned a ”default” GP, based on their residential address.

The number of authorized family doctor positions allocated to each municipality is regulated by the Norwegian Directorate of Health. Hence, the family doctor market is not characterized by free entry, and the degree of competition is limited. As we show below, a consequence of this system is that the majority of family doctors are normally subjected to excess demand, implying that vacant patient slots are filled immediately.

The data we use in the present paper are collected from Norwegian administrative registers. They cover all workers and physicians in Norway during the period from January 2002 through December 2010, with individual level (encrypted) monthly data on person-GP-linkages. For each worker, we have monthly information on physician-certified absence, as well as on gender, age, and place of residence. For each GP, we have information on gender and age, the actual and desired number of patients, and their type of remuneration. The latter is an indicator variable denoting whether the physician is running his/her own business (denoted variable wage) or is employed by the municipality in a fixed-wage contract (denoted fixed wage).

Using these data, we construct for each employee, a monthly variable indicating the number of physician-certified absence days, adjusted for grading.[3] In addition, we construct for all persons aged 18-62, a monthly indicator giving the (encrypted) identity of the chosen panel doctor.

3 Identifying and estimating GP leniency

We assume that for each GP and for each point in time, there exists a latent variable characterizing the degree of gatekeeper leniency; i.e., the readiness to certify the use of public funds in cases where there is scope for subjective judgment. Systematic differences in leniency may reflect genuine differences in the assessment of what are the "right" certification decisions (from a purely medical or from a social cost-benefit point of view), as well as differences in the willingness to deviate from these right decisions in order to satisfy the patients. The latter may (or may not) depend on the physician's competitive position, as reflected in the demand for his/her services.[4]

Although gatekeeper leniency is relevant for several services provided by GPs – such as referrals to specialists, prescription of subsidized drugs, and certification of disability insurances – we identify and estimate it solely on the basis of sick pay certification in this paper.[5] There is an existing literature indicating that many GPs find their role as sick pay certifier problematic, and that they sometimes are concerned that they could lose customers if they refuse such requests; see, e.g., Carlsen and Norheim (2005), Nilsen et al. (2011), and Winde et al., (2012). There is also evidence showing that GPs rarely overrule their patients' requests for sick leave certificates (Englund and Svärdsudd, 2000; Carlsen and Nyborg, 2009). However, even though it might be difficult outright to reject a request for sick leave, the GP may behave in a way that makes it difficult for the patient to press the issue, and also play an important role as a persuader

[3] It is common to use graded absence certificates in Norway, implying that the worker is only partially absent from the workplace, see Markussen et al. (2012) for details. If, for example, a worker has a 50% absence spell for 10 days, we count this as 5 days of absence.

[4] While we focus on the influence of self-interest here, there are also other reasons why physicians deviate from a principal's understanding of the appropriate gatekeeping practices. Many physicians view themselves as advocates of their patients; see, e.g., Schwartz (2002). The World Medical Association's Code of Medical Ethics, states that "a physician shall owe his/her patients complete loyalty and all the scientific resources available to him/her" (World Medical Association, 2006), and the Charter of Medical Professionalism states patient welfare as the first fundamental principle: "Market forces, social pressures, and administrative exigencies musts not compromise this principle" (Medical Professionalism Project, 2002, p. 244).

[5] For the period considered in this paper, we do not have access to data that could facilitate the use of other gatekeeping decisions to identify and estimate GP leniency.

and motivator. On the basis of exogenous GP switches caused by transfers of complete customer lists, Markussen et al. (2013) report that a change of GP has considerable impact on the workers' absence behavior.

We compute a monthly family doctor leniency indicator based on the overall certification of sick pay days for employed customers, controlled for customer composition and calendar time. This implies that we attribute to the family doctor all absences certified for his/her patients, even when they are certified by other doctors. The main reason for this is that the number of absence certificates issued in the family doctor's own name is heavily influenced by opening hours. Many GPs combine the family doctor business with work at hospitals or elderly care institutions 1-2 days a week or run part-time businesses for other reasons, and by attributing all absence certificates to a worker's family doctor we ensure that variations in availability are not falsely interpreted as variations in gatekeeping leniency. An additional rationale behind this "intention to treat" strategy is that the family physician is likely to have an influence on patients' absence behavior well beyond his/her own issuing of absence certificates; e.g., through communication with colleagues at the medical center and through his/her referral practice.

For each family physician in Norway, we estimate the degree of gatekeeper leniency – month by month – on the basis of the employed customers' sick leave certificates. This raises a tricky identification problem: How can we properly isolate the GPs' leniency behavior from differences in the demand for sick leave driven by customer composition? Our primary strategy for solving this problem is to control for worker fixed effects as well as worker age (the latter on an annual basis). This strategy ensures that all time-invariant and age-related differences between the workers' demand for sick leave certificates are controlled for, and that GP leniency is identified on the basis of *changes* in customer composition only. However, this approach will not ensure identification if there is a systematic association between health shocks and GP choice, e.g., such that some GPs tend to recruit workers who have been exposed to particularly negative health shocks. To address this concern, we also employ an alternative strategy where we identify GP leniency on the basis of a *predicted* list of customers, thereby blocking the influence of endogenous GP choices.

Table 1. Workers and GPs in Norway. January 2002 - December 2010

(a) Workers	
Number of workers	2,480,466
Average number of monthly worker observations	63.95
Distribution of prescribed absence days over workers (whole observation period)	
Mean	114.54
Percent with positive absence	74.4 %
Distribution conditional on positive	
Mean	153.96
10^{th} percentile	8
Median	75
90^{th} percentile	413
Distribution of prescribed absence days over worker-months	
Mean	1.79
Percent with positive absence	9.0 %
Distribution conditional on positive	
Mean (standard deviation)	19.81
10^{th} percentile	4
Median	15
90^{th} percentile	38
(b) Physicians	
Number of physicians	6,782
Average number of months in practice during our data window	62.5
(c) Worker-physician relationships	
Average GP turnover (percent of workers shifting GP during a month)	1.06
Average patient list worker turnover (percent of employed list members that were not employed or list-members in previous month)	3.72

Note: The distribution of prescribed absence days over worker-months refer to the number of days certified in each month – not the number of days actually absent. Hence, these numbers may exceed 31 days

These exercises are based on a total number of 2.5 million workers over a period of nine years, yielding as much as 158 million monthly observations divided between around 6,800 physicians.[6] Table 1 presents some descriptive statistics for workers (panel a), physicians (panel b) and worker-physician relationships (panel c). The description of worker absenteeism in panel (a) shows that around three quarters of the workers have at least some physician-certified absence during our estimation period. The distribution is skewed, however, and among those with

[6] Note that approximately 20 % of the 6,782 physicians used to estimate GP leniency are not "genuine" family doctors; i.e., we cannot identify a specific person connected to the practice. This happens, for example, when a patient list is operated by more than one physician and in cases where the regular GP has a temporary substitute. Although we are not going to use the estimated physician leniency indicators for these "non-GPs" later in this paper, we include them in the analysis here, since their customers help identifying the worker-fixed effects and thus indirectly the leniency of other physicians. Note also that the physician-identity used in this paper changes in the very rare cases where a physician moves his/her practice from one municipality to another.

positive absence the mean is roughly twice as large as the median. Figure 1 provides a more comprehensive picture of the overall distribution of the absence days in our data over the worker population. It shows that the one percent most absence-prone workers accounts for approximately 10 % of overall absenteeism, whereas the 10 % most absence-prone workers accounts for approximately 40 %. Still, most workers do have a certified absence spell from time to time; hence the foundation for identification of GP leniency rests on a large number of workers. And, as shown in panel (c), there are considerable movements of workers between physicians. Every month, approximately one percent of the workers shift family doctor, and for each physician the monthly turnover rate of employed patients (i.e., patients used to identify their sick leave certification practice) is close to four percent.

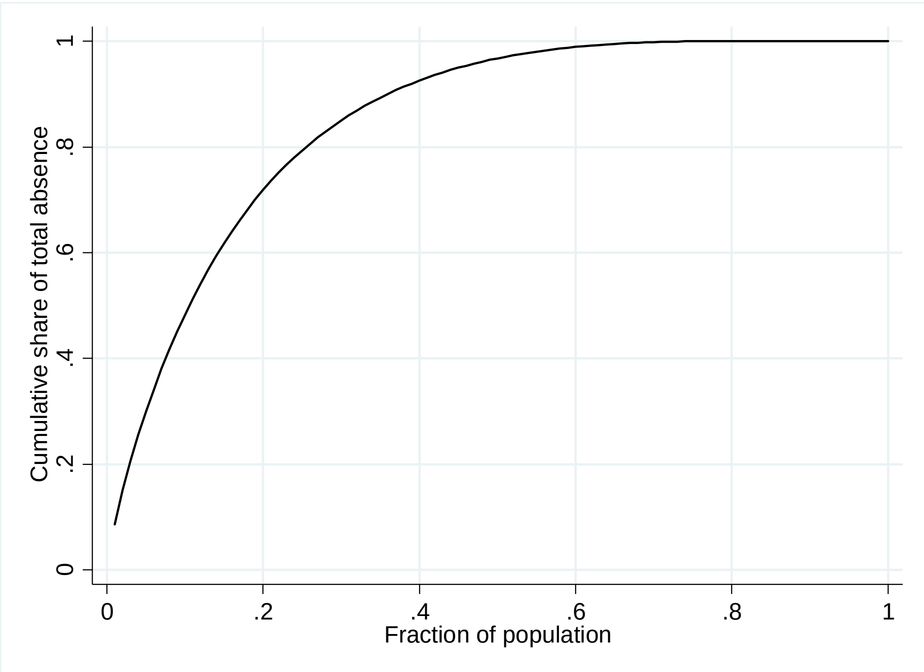

Figure 1. The dispersion of physician-certified absence days in Norway 2002.1-2009.12 among all workers represented in the dataset.

3.1 A leniency index based on actual GP-worker affiliations

To illustrate our primary identification strategy, let y_{it} be the number of grade-adjusted sick pay days prescribed for worker *i* in calendar month *t*.[7] We set up the following regression equation:

[7] Note that this is not the same as the number of certified sick pay days in the same month, as we attribute all absence days to the month they were *certified*. As the typical length of a single certification period is around 2-3 weeks (19 days on average), these periods will often stretch into the next month.

$$y_{ijt} = \alpha_i + \delta_a + \mu_{jt}^* + e_{ijt} \ , \tag{1}$$

where α_i is a worker fixed effect, δ_a is an age fixed effect, μ_{jt}^* is a physician-by-month fixed effect, and e_{ijt} is a residual. We may think of the worker fixed effects as representing the *demand* for sick leave certificates, whereas the physician-by-month fixed effects represent the *supply.*

We estimate Equation (1) with a least squares algorithm developed by Gaure (2013) and take out the resultant $\hat{\mu}_{jt}^*$.[8] As an indicator for physician *j*'s degree of gatekeeper leniency in month *t,* we then define

$$\hat{\mu}_{jt} = \hat{\mu}_{jt}^* - \sum_{k=1}^{J} \omega_{kt} \hat{\mu}_{kt}^* \ , \tag{2}$$

where ω_{kt} *is* the fraction of workers affiliated to physician *k* in month *t.* This implies that a physician's leniency in month *t* is computed as his/her month *t* fixed effect's deviation from the contemporary weighted mean of physician fixed effects. Hence, we disregard any common time trends in physician leniency. The reason for this is that we cannot separately identify common time trends in physician leniency from other time developments, related to, e.g., the population's health or work norms and fluctuations in epidemics.

Note that within-worker fluctuations in the demand for sick leave certificates induce a simultaneity problem in Equation (1), as the residuals become correlated with the physician-by-month fixed effects. This represents a fundamental identification problem in the sense that we are unable to separate the GPs monthly leniency behavior from joint fluctuations in his/her customers' demand for sick leave due to, say, correlated health shocks. We will take this simultaneity into account when we examine the workers' choice of GP in the next section by exclusively focusing on choice situations where the worker has not contributed to the computation of the leniency indicators for the GPs entering his/her own choice set. To reduce the noise generated by fluctuation in other customers' health, we will also aggregate the monthly index over 12-month intervals.

To check whether these 12-month averages properly represent physician behavior as opposed to fluctuations in patient health, we assess in Figure 2 their stability over time by providing

[8] The adjusted R-square from this regression is 0.10.

cross plots of the physicians' estimated 12-month-averaged leniency indicators computed for different time periods. While panel (a) compares leniency indicators calculated on the basis of adjacent calendar years, panels (b) and (c) compare indicators calculated with 4 and 8 year distances, respectively. It is clear that the leniency indicators averaged over different time periods are highly correlated, even when there are several years between the measurement periods. If our leniency indicators had primarily reflected random fluctuations in the workers' demand for sick leave certificates, there would be no reason to expect such a pattern.

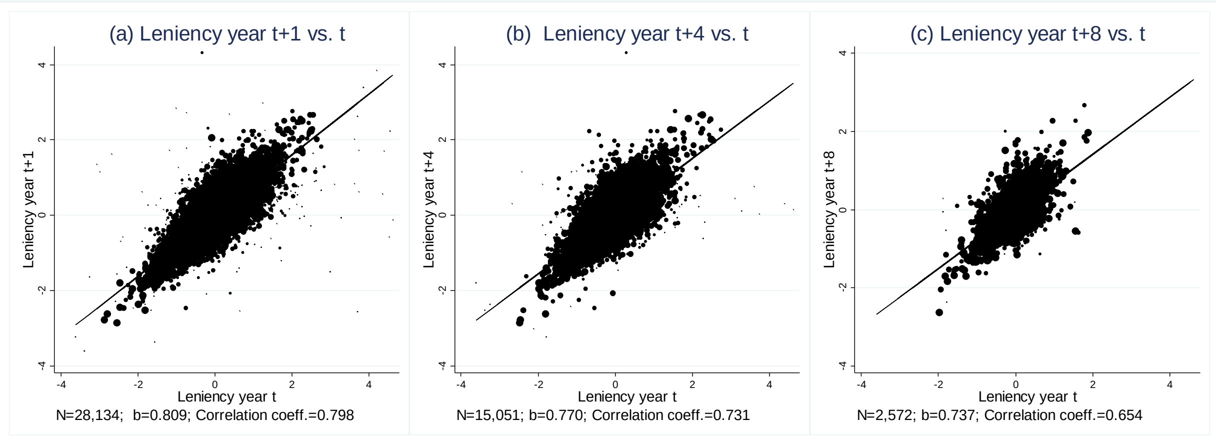

Figure 2. Cross-plots of estimated physician leniency indicators averaged over different 12-month periods.

Note: The sizes of the dots reflect the number of worker-months behind each average. The numbers of observations, the point estimate behind the illustrated regression lines, and the correlation coefficients are reported below each panel.

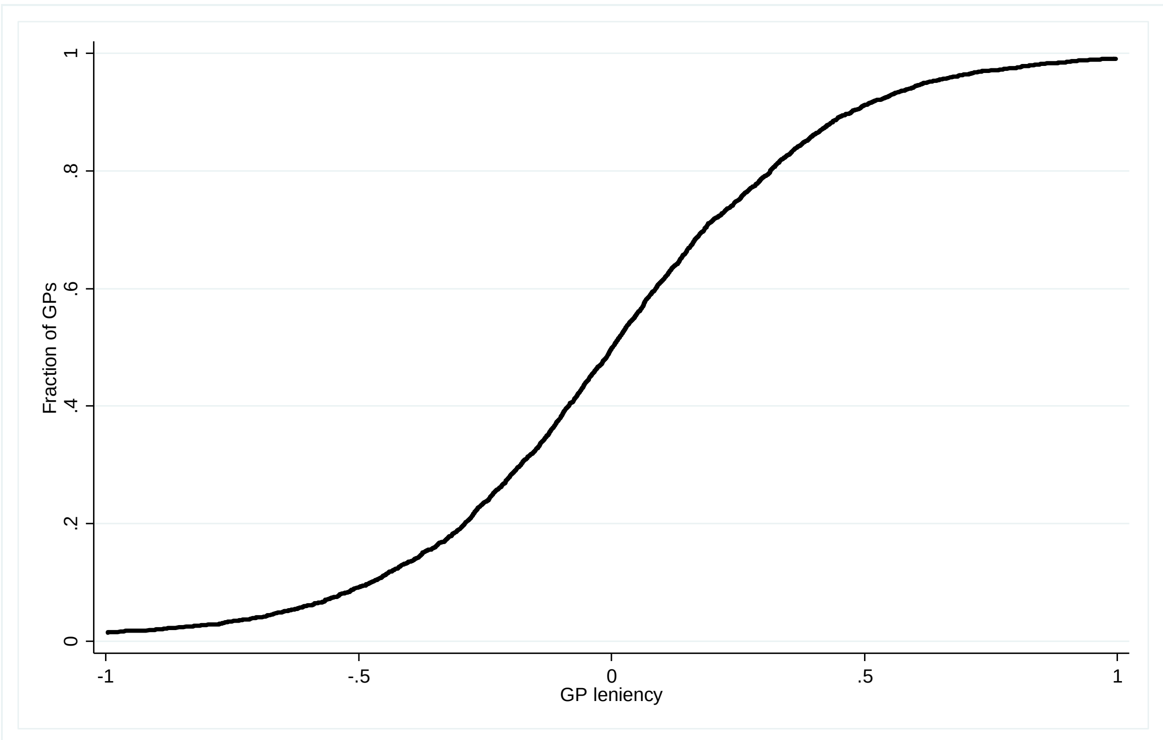

Figure 3. The distribution of time-averaged leniency indicators for the physicians in Norway.

Note: The distribution function in panel is based on the weighted distribution of the time-averaged leniency indicators (with number of employed customers used as weights).

Assuming that fluctuations in the demand for sick leave are randomly distributed across GPs conditional on the worker fixed effects, we can illustrate the implied overall difference in gatekeeper leniency between GPs by averaging over all monthly observations of $\hat{\mu}_{jt}$ available for each physician j. The resultant distribution of time-averaged leniency indicators (weighted by the overall number worker-months) is presented in Figure 3. It shows that as we move from the strictest to the most lenient GP, the expected number of prescribed absence days during a month increases by around 2, ceteris paribus. Given that the average number of prescribed absence days is around 1.8 (see Table 1), this is a considerable degree of variation.

3.2 A leniency index based on predicted GP-worker affiliations

As explained above, it is possible that the within-worker fluctuations in the demand for sick leave (e.g., due to health shocks) are not randomly distributed with respect to their choices of GPs. If some GPs disproportionally recruit workers recently exposed to adverse health shocks, their estimated leniency derived from Equations (1) and (2) will be biased upwards. To address this concern, we also compute a leniency index based on the GPs' *predicted* rather than *actual* customers. In the spirit of Gowrisankaran and Town (1991), we use geographical distance between worker and GP as the key prediction variable, together with similarity in age and equal sex. More specifically, we allocate each worker to the GP operating closest to the place of residence. If there is more than one GP with the same geographical distance, we assign the worker to a GP of the same sex, and if there is more than one of the same sex, we choose the one with the smallest difference in age.[9] If there is still more than one alternative remaining, we choose the predicted GP by a random draw. Based on the resultant monthly predicted GP-worker affiliations, we re-estimate Equations (1) and (2), but this time without worker-fixed effects (since there is not sufficient time variation in the predicted GP-worker affiliations to facilitate identification of these effects), but including controls for the workers' sex and age. Unfortunately, given the large number of GPs available to most workers, our prediction model is not able to allocate more than around 14 % of the workers to their correct GP; hence, there is a lot of attenuating noise in the resultant leniency indicator. Table 2 provides a comparison of the actual and the predicted GP-patient matches and the associated distributions of estimated GP leniency. The prediction model clearly assigns GPs that, compared to the actual assignments,

[9] We show in Section 4 below that these factors are indeed important for workers' selection of GP.

are closer to the customers in terms of geography, gender, and age. However, the overall variation in estimated GP leniency is almost exactly the same, and the distributions of leniency estimates appear similar. The correlation between the two indicators is 0.17. Although this is not very impressive, it suggests that the simple prediction model is likely to pick up systematic differences between GPs. Hence, we will use it as part of a sensitivity analysis in the next section

Table 2. Characteristics of actual and predicted GP-worker matches and the associated leniency estimates

	Actual GP-worker match	Predicted GP-worker match
Correct matches (fraction)	1	0.14
GP in own local area (fraction)	0.41	0.81
GP within walking distance (fraction)	0.26	0.50
Mean travel distance to GP (minutes by car)	24.2	11.4
Mean GP age (years)	48.4	45.8
GP female (fraction)	0.28	0.37
GP and patient same gender (fraction)	0.61	0.88
GP and patient of similar age (+/- 5 years) (fraction)	0.30	0.52
Variation in estimated GP leniency (standard deviation of GP leniency estimates)	0.45	0.45
1st percentile	-1.15	-0.89
25th percentile	-0.27	-0.28
75th percentile	0.27	0.28
99th percentile	1.14	1.26
Correlation between estimated leniency based on actual and predicted GP-patient match	0.17	

4 The workers' choice of family doctor

In this section, we examine the degree to which workers take the gatekeeper leniency reputation (as reflected in actual sick-leave prescription for other workers) into account when they choose their family doctor? In the main part of the analysis we condition on a physician shift taking place; i.e., we examine situations where a worker chooses a (for him/her) new family doctor. This way, we zoom in on situations where we know that *active* GP choices are made, and we avoid by construction the identification problem that arises when persons making these choices have themselves contributed with data used to estimate the degree of leniency for the GPs entering their choice set. We return to an analysis of the decision to deselect an existing family doctor towards the end of the section, using a simplified leniency indicator based on the number of absence days recently prescribed by the same GP for *other* workers.

4.1 Data and model

Since GP leniency is intrinsically a latent characteristic, workers must obviously choose GP with imperfect information about their gatekeeping-practices. It seems probable, though, that there exist local rumors regarding physicians' practice styles, which at least to some degree mirror their true leniency.[10] We will assume adaptive behavior in the formation of a GP's leniency reputation. More specifically, in a baseline model we assume that the local reputation corresponds to the observed degree of sick pay-prescription (adjusted for patient composition) as computed in the previous section, but averaged over the past 12 months. To ensure that workers choosing a new family doctor have not themselves contributed to the computation of the leniency index for any of the GPs entering the choice set, we also require that the "old" family doctor was retained for at least 12 months. In separate robustness analyses below, we instead assume that reputations are based on the GPs' behaviors over the past 6 or 18 months (with a corresponding requirement regarding the length of the old family doctor attachment).

The starting point of our analysis is the set of all family doctor shifts that took place in our data period, and the endogenous outcome is the identity of the new family doctor. We drop from the analysis shifts that took place as part of a "mass-transfer" where a whole (or a large fraction of) a list was transferred (sold) to another physician, since these shifts provide no information about individual preferences.[11] For each shift situation we try to identify the worker's choice set, i.e., the set of GPs that made up the menu of feasible choices. To do this, we first record for each worker the correct local area of residence.[12] We then identify the set of potential GP choices by assuming that all physicians who have at least 10 other patients from the same local area belong to the choice set. In separate robustness analyses, we instead use thresholds of 5 or 50 patients to identify the choice set.

[10] There is now a web-portal (www.legelisten.no) designed to help citizens choose family doctor, where customers share their views and experiences with named physicians, often with a focus on their gatekeeping practices. This portal did not exist in our data period (it was established in 2012), but similar (local) discussion groups were probably common. Note that since the data used in the present paper are encrypted (due to privacy protection considerations), we are not able to compare our estimated leniency indicators to (later) customer views expressed at the web-portal.

[11] If more than 10 customers make exactly the same shift at exactly the same time, we interpret them as resulting from a mass-shift, and drop them from the analysis. In practice, these mass-shifts typically involve a much larger number of customers, and if we set the threshold to 50 instead of 10, we get almost exactly the same results as those reported here.

[12] In this exercise, we use "local areas" corresponding to the statistical tracts ("delområder"), drawn up by Statistics Norway; see Statistics Norway (1999) for details. They are designed to encompass neighborhoods that naturally interact, e.g., by sharing common service/shopping center facilities. There are 1,535 local areas in Norway, and a typical local area comprises around 3,100 inhabitants.

One common reason for changing family doctor is relocation to a new local area. Since we have data on residential local area on an annual basis only, we are in some cases not able to identify with certainty the correct GP choice set. In these cases we construct choice sets corresponding to both the previous and the new address, and we then include in the analysis the choice set in which we see that a GP was actually selected. If the selected GP enters both choice sets, we randomly select one of them.[13]

In total, we identify around 671,000 active choice situations satisfying our requirements. However, there are several cases where workers choose GPs that were not assumed by us to be part of their choice set (this can happen, for example, if they chose a GP close to where they work rather than close to where they live). After having dropped these choice situations from the analysis, we end up with 394,720 active GP choices that can be used in our statistical analysis. Table 3 gives an overview of the size distribution of the resultant choice sets. On average, a worker can choose between 74.5 different physicians, but the variation is large – from less than 10 alternatives in some rural areas to more than 300 in some big city districts.

Table 3. Cases of active choices of family doctor in Norway. January 2003 - December 2009

Total number of choice situations	394,720
Average size of choice set (number of available physicians)	74,5
smallest	2
10^{th} percentile	9
25^{th} percentile	20
median	44
75^{th} percentile	118
90^{th} percentile	172
largest	321

Note: Compared to the data used in the previous section, the data window is cut by 12 months at each end. The first 12 months is lost due to our condition of at least 12 months attachment to the previous/deselected GP. The last 12 months is lost for the reason that we need "next year's" records to check for migration to a new local area in order to identify the correct GP choice set.

Based on these choice sets, we set up a model designed to explain actual GP choices. We do this on the basis of GP characteristics within the framework of a conditional logit model (McFadden, 1974), similar to the strategy used by Santos et al. (2017) to study the impacts of clinical quality on the choice of family doctors in England. We assume that worker i's utility associated with choosing a particular GP j can be expressed as

[13] If we instead drop these observations from the analysis, the results are almost exactly the same as those reported below.

$$U_{ij} = \mathbf{x_{ij}\beta} + v_{ij}, \tag{3}$$

where $\mathbf{x_{ij}}$ is a covariate vector that differs across the alternative physicians and v_{ij} are unobservable taste components. Assuming that the v_{ij} are independently distributed according to the type I extreme value distribution, we can write the choice probabilities as

$$P(d_i = j) = \frac{\exp(\mathbf{x_{ij}\beta})}{\sum_{h \in J_i} \exp(\mathbf{x_{ih}\beta})}. \tag{4}$$

where $d_i = j$ if physician j was worker i's chosen alternative from his/her choice set J_i. This model specification rests on the assumption of "independence of irrelevant alternatives" (IIA); i.e., that the relative choice probabilities for any two alternative physicians depend only on the attributes of those two.

The explanatory variable of main interest is the physicians' reputations on gatekeeper leniency, as reflected in $\hat{\mu}_{jt}$ averaged over the past 12 months; i.e., for a GP choice made in month t we include as a reputation indicator for physician $j \in J_i$ the variable

$$\hat{\mu}_{jt}^{S} = \tfrac{1}{S} \sum_{s=t-S}^{t-1} \hat{\mu}_{js}, \tag{5}$$

with $S=12$ and $\hat{\mu}_{jt}$ defined in Equation (2). In addition, $\mathbf{x_{ij}}$ contains covariates describing the geographical distance between the worker's and the physician's locations, whether the GP offers an emergency service or not, the GP's gender and age, and the existence of available patient slots at the start of the month.[14]

4.2 Main results

The estimation results are shown in Table 4, for the full dataset of all workers (column (1)) as well as for men and women and for two different age groups separately (columns (2)-(5)). As

[14] Since we only have precise geographical information on the location of workers' homes, we compute the distance-to-physician variables by assuming that each GP is located in the business center of the local area in which the largest number of his/her patients live. With the aid of geographical positioning data, we compute the distance between the worker's home and each physician's office. We then define two variables; one indicator variable for "walking distance" (less than 1,500 meters) and one scalar variable indicating the number of minutes it will take to drive by car.

noted above, the IIA assumption implies that the relative choice probabilities for any two alternative physicians depend only on the attributes of those two physicians. Thus for any alternatives $j,k,$ it follows from (4) that we have

$$\frac{P(d_i = j)}{P(d_i = k)} = \frac{\exp(\mathbf{x}_{\mathbf{ij}}\boldsymbol{\beta})}{\exp(\mathbf{x}_{\mathbf{ik}}\boldsymbol{\beta})} = \exp\left[(\mathbf{x}_{\mathbf{ij}} - \mathbf{x}_{\mathbf{ik}})\boldsymbol{\beta}\right]. \tag{6}$$

This implies that we can interpret the exponentiated coefficients as the estimated change in the relative choice probabilities associated with a marginal change in the corresponding variable. The coefficient estimate associated with the leniency variable $\hat{\mu}_{jt}^{12}$ of 0.13 in column (1) thus implies that when a physician's leniency reputation increases with one unit – implying one additional expected day of certified absence per month, ceteris paribus – the relative probability of being chosen over each of the competing GP alternatives increases by $100 \times (\exp(0.13) - 1) = 13.9$ %.

Table 4. Workers' choice of new GP. Estimation results for baseline model (standard errors in parentheses)

	(1) All	(2) Women	(3) Men	(4) Age 18-44	(5) Age 45-66
GP leniency	0.134*** (0.004)	0.136*** (0.006)	0.130*** (0.006)	0.117*** (0.005)	0.197*** (0.009)
Within walking distance	0.581*** (0.006)	0.565*** (0.009)	0.595*** (0.009)	0.580*** (0.007)	0.548*** (0.015)
Estimated driving time if not walking distance (minutes)	-0.108*** (0.000)	-0.103*** (0.001)	-0.111*** (0.001)	-0.110*** (0.001)	-0.099*** (0.001)
Emergency service provided	0.013*** (0.004)	0.031*** (0.006)	-0.001 (0.005)	0.017*** (0.004)	0.021** (0.009)
Same sex GP	0.664*** (0.004)	0.532*** (0.005)	0.737*** (0.006)	0.669*** (0.004)	0.652*** (0.009)
Female GP	-0.099*** (0.004)			-0.068*** (0.005)	-0.220*** (0.009)
GP age	-0.017*** (0.000)	-0.021*** (0.000)	-0.014*** (0.000)	-0.016*** (0.000)	-0.022*** (0.001)
GP younger than patient	0.014* (0.008)	-0.005 (0.011)	0.029*** (0.011)	0.103*** (0.014)	-0.034** (0.014)
GP older than patient	-0.139*** (0.006)	-0.110*** (0.009)	-0.166*** (0.009)	-0.178*** (0.007)	0.029** (0.014)
Free capacity at start of month	0.985*** (0.004)	0.884*** (0.006)	1.078*** (0.005)	1.005*** (0.004)	0.900*** (0.0086)
N #choices×size of choice set	29,742,433	14,048,568	15,693,865	24,611,473	5,130,960

Note: GP leniency is defined in Eq. (5). Walking distance is defined as less than 1,500 meters (0.9 miles). Estimated driving times are computed from "open street map". ***(**)(*) indicates statistical significance at the 1(5)(10) % levels, respectively, based on reported standard errors. For the effect of GP leniency, these standard errors are slightly underrated, as this explanatory variable is itself estimated with uncertainty. A Monte Carlo Analysis indicates that the true standard errors in row 1 are approximately 15 % larger than those reported.

Note that the standard errors reported in this section are a bit smaller than the true variability of the point estimates, since the key variable (GP-leniency) is generated from a separate estimation and thus also subjected to statistical uncertainty. To assess the impact of this phenomenon, we have performed a full non-parametric bootstrap for a reduced version of the model.[15] The results of this exercise indicate that the correct standard errors for the estimated impacts of GP leniency are approximately 15 % larger than the standard errors reported in the table. Given the high precision in these estimates, this is of little practical importance, and has clearly no consequences for any of the significance statements.

To put the estimated effects of leniency reputation into perspective, we report in Table 4 also the estimated impacts of other physician characteristics. Geographical proximity to the worker is clearly of great importance; and being located within walking distance of the worker raises the relative choice probability considerably. For example, substituting walking distance for a 10 minutes' drive raises the relative choice probability by a factor of $\exp(0.58+10\times0.11)=5.37$. Workers also appear to prefer doctors with emergency service. Apart from these GP practice characteristics, the results in column (1) show that workers care about the GP's gender and age. They prefer GPs of the same sex as themselves. Workers also prefer GPs that are either of the same age as themselves or younger. Finally, the table shows that having vacant slots (free capacity) at the start of the month raises the relative probability of being chosen by a factor around 2.7. In principle it is not possible to choose a doctor without vacant capacity (see Section 2); hence, it could be argued that GPs without vacant capacity should not belong to the choice set at all. However, as we show in the next section, many GPs accept new customers despite having exceeded their desired list lengths, and high turnover implies that many patient slots become vacant during the course of a month. In a robustness exercise, we drop physicians without vacant capacity from the choice set.

It may be of some interest to examine how the weight attached to GP leniency vary across demographic groups, and columns (2)-(5) in Table 4 report results for women and men and for young and old workers separately. The estimated effect of GP leniency is very similar across these groups, however, with the exception that older workers seem to give higher priority to the GP's leniency reputation than younger workers do. This is a natural consequence of the fact that older workers generally have a higher demand for both sick leave certificates and other

[15] The bootstrap described here requires large computational resources; hence it has not been feasible for us to use this strategy for all reported standard errors in the paper. In this section, we have used the model reported in Table 5, Column (2) for this particular purpose, since it contains considerably fewer observations than the other models, which has made it feasible to implement a sufficient number of trials. More specifically, our assessment of the true standard errors is based on 104 bootstrap trials, with each trial involving the following two steps: First, we resample (with replacement) the 158 million worker-months and estimate the leniency indicator. Then we resample (again with replacement) persons about to select a new GP (including their choice set) and estimate the GP choice model.

rationed services provided by the GP, such as specialist referrals and prescription of subsidized drugs.

Table 5. Workers' choice of new GP. Estimation results for baseline model (standard errors in parentheses)

	(1) All	(2) Absent t-1	(3) Present t-1	(4) Present up to GP choice	(5) Always present
GP leniency	0.134*** (0.004)	0.285*** (0.010)	0.100*** (0.0046)	0.051*** (0.008)	0.031*** (0.009)
Within walking distance	0.581*** (0.006)	0.527*** (0.015)	0.587*** (0.007)	0.649*** (0.012)	0.632*** (0.014)
Estimated driving time if not walking distance (minutes)	-0.108*** (0.000)	-0.099*** (0.001)	-0.110*** (0.001)	-0.114*** (0.001)	-0.120*** (0.001)
Emergency service provided	0.013*** (0.004)	0.035*** (0.009)	0.009** (0.004)	-0.039*** (0.007)	-0.048*** (0.008)
Same sex GP	0.664*** (0.004)	0.556*** (0.010)	0.691*** (0.004)	0.725*** (0.007)	0.706*** (0.009)
Female GP	-0.099*** (0.004)	-0.234*** (0.010)	-0.067*** (0.005)	-0.036*** (0.008)	-0.052*** (0.009)
GP age	-0.017*** (0.000)	-0.022*** (0.001)	-0.016*** (0.000)	-0.012*** (0.000)	-0.011*** (0.001)
GP younger than patient	0.014* (0.008)	-0.065*** (0.018)	0.033*** (0.009)	0.062*** (0.016)	0.094*** (0.019)
GP older than patient	-0.139*** (0.006)	-0.096*** (0.015)	-0.149*** (0.0069)	-0.160*** (0.012)	-0.158*** (0.014)
Free capacity at start of month	0.985*** (0.004)	1.018*** (0.009)	0.978*** (0.004)	1.166*** (0.007)	1.166*** (0.008)
N #choices×size of choice set	29,742,433	5,280,516	24,461,917	9,785,283	7,183,834

Note: GP leniency is defined in Eq. (5). Walking distance is defined as less than 1,500 meters (0.9 miles). Estimated driving times are computed from “open street map”. ***(**)(*) indicates statistical significance at the 1(5)(10) % levels, respectively; see also note to Table 4.

It appears probable that workers who are on sick leave around the time of GP selection are more concerned about a prospective family doctors leniency-reputation than workers who are not on sick leave. In Table 5, we report separate estimation results for worker groups distinguished by their absence behavior. Column (1) first repeat the estimation results for the complete dataset, whereas column (2) report results for the subpopulation characterized by absence in the month just before the GP choice was made. As expected, the GP’s leniency reputation is much more important for the workers with a sick leave certificate issued in the month prior to the shift of GP than for workers who were present in that month (column 3). Moving on to workers who have not been absent at all (in our data period) prior to the choice of GP (column 4), the estimated effect becomes even smaller. And finally, looking at workers who are non-absent throughout our estimation period, the estimated effect declines further toward zero. While this pattern of increasing weight on GP leniency the higher is the worker's absence propensity is interesting in its own right, it also serves as a confirmation that our estimates really represents causality. Had our model captured some spurious relationship between leniency and GP choice, there would be no reason to expect such a systematic pattern

4.3 Robustness

The results presented so far build on functional form assumptions and definitional choices for which we have no clear evidence-based guidance. One may worry that these choices have been made on the basis of a data-mining exercise seeking to identify the most "publishable" results. To allay such concerns, we present in Table 6 some results based on alternative assumptions. Columns (2)-(4) show results based on alternative assumptions regarding the workers' choice sets. In column (2), we have assumed that all physicians that are chosen by at least five other individuals in a local area belong to a worker's choice set, instead of 10, as in the baseline model (recall that a local area on average consists of 3,100 inhabitants). This raises the average size of the choice sets from 74.5 in the baseline model to as much as 108.8. Yet the estimated coefficient on GP leniency remains unchanged. In column (3), we have instead restricted the choice sets to physicians chosen by more than 50 other individuals in the local area. This reduces the average size of the choice sets to 22.7 physicians. The estimated effect of GP leniency then becomes a bit smaller, although it is still highly significant. In column (4), we maintain the threshold of 10 employed customers in the local area, but remove from the choice set all physicians without vacant capacity at the start of the month. This reduces the average size of the choice set to 37.4 physicians. And the estimated effect of leniency becomes a bit larger than in the baseline model.

In columns (5) and (6), we present results based on alternative assumptions regarding the formation of GP reputation. In column (5), we have included only six months of past leniency indexes to compute average leniency; i.e., we have substituted $\hat{\mu}_{jt}^{6}$ for $\hat{\mu}_{jt}^{12}$; see Equation (5). This also implies that we raise the number of observations by as much as 20 %, since we in this model can reduce the conditioning period for having been affiliated to the same GP from 12 to 6 months. Yet, the estimated effect of GP leniency remains the same. In column (6), we instead include as much as 18 months of past leniency indexes to compute average leniency (using $\hat{\mu}_{jt}^{18}$), and consequently throw out around 20 % of the observations. Still, the estimated effect of GP leniency remains almost unchanged.

Table 6. Workers' choice of new GP. Robustness (standard errors in parentheses)

	(1) Baseline	(2) >4 patients	(3) >49 patients	(4) Vacant	(5) 6 months	(6) 18 months
GP leniency	0.134*** (0.004)	0.136*** (0.004)	0.086*** (0.005)	0.190*** (0.006)	0.147*** (0.003)	0.121*** (0.005)
Within Walking distance	0.581*** (0.006)	0.642*** (0.006)	0.505*** (0.008)	0.627*** (0.008)	0.580*** (0.006)	0.582*** (0.007)
Estimated driving time if not walking distance (minutes)	-0.108*** (0.000)	-0.118*** (0.000)	-0.053*** (0.001)	-0.109*** (0.001)	-0.104*** (0.000)	-0.110*** (0.001)
Emergency service provided	0.013*** (0.004)	-0.0050 (0.0038)	-0.003 (0.005)	0.026*** (0.005)	0.023*** (0.004)	0.012*** (0.004)
Same sex GP	0.664*** (0.004)	0.667*** (0.004)	0.645*** (0.004)	0.674*** (0.005)	0.657*** (0.004)	0.668*** (0.004)
Female GP	-0.099*** (0.004)	-0.093*** (0.004)	-0.112*** (0.005)	0.036*** (0.006)	-0.069*** (0.004)	-0.117*** (0.005)
GP age	-0.017*** (0.000)	-0.018*** (0.000)	-0.016*** (0.000)	-0.019*** (0.000)	-0.020*** (0.000)	-0.015*** (0.000)
GP younger than patient	0.014* (0.008)	0.008 (0.008)	0.0342*** (0.009)	-0.011 (0.010)	0.019*** (0.007)	0.012 (0.009)
GP older than patient	-0.139*** (0.006)	-0.134*** (0.006)	-0.149*** (0.007)	-0.118*** (0.008)	-0.147*** (0.006)	-0.136*** (0.007)
Free capacity at start of month	0.985*** (0.004)	0.979*** (0.004)	0.989*** (0.004)		1.038*** (0.004)	0.954*** (0.004)
N (#choices×size of choice set)	29,742,433	44,978,235	7,177,562	10,502,604	36,747,433	24,256,401

Note: Columns (2) and (3) are based on choice sets identified on the basis of at least 5 or 50 other individuals in the local area having selected the physician in question. Column (3) is based on at least 10 persons having selected the physician (as in the baseline model), but only with physicians with vacant capacity at the start of the month included. Columns (5) and (6) are based on alternative definitions of GP reputation, based on the last 6 or 18 months, respectively. ***(**)(*) indicates statistical significance at the 1(5)(10) % levels, respectively; see also note to Table 4.

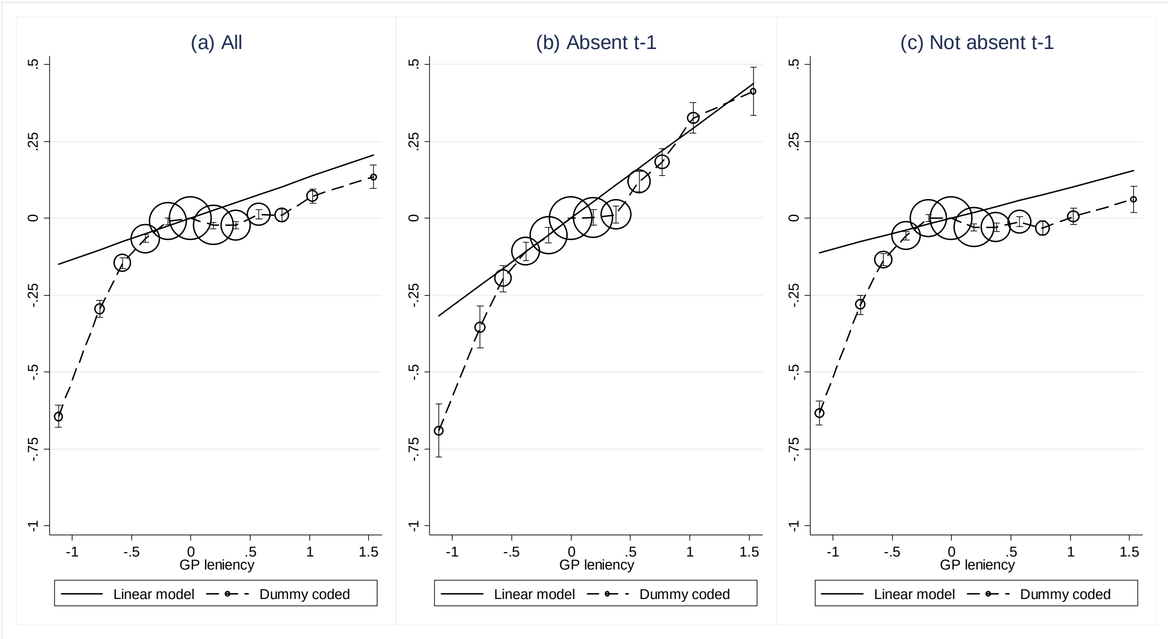

Figure 4. Non-linear estimates of the impacts of leniency
Note: The leniency indicator variables are defined such that each of the 10 bins on the horizontal axis have exactly the same length (0.2 days), except at the two ends where each category contains all leniency observations below the 1st and above the 99th percentiles in the leniency distribution, respectively. The regression lines are based on the results in Table 5, column (1)-(3)

An additional concern could be that we have used a rather restrictive functional form assumption for the relationship between GP leniency and customer utility. There is of course no behavioral theory that can be called upon to justify a specific functional form in this case, and the linearity assumption must be viewed primarily as a choice motivated by convenience and interpretability. As an alternative, we have therefore estimated models where we have dummy-coded the GP-leniency indicator into 10 brackets. The result from this exercise is illustrated in Figure 4, where we have plotted the estimated coefficients for the models corresponding to columns (1)-(3) in Table 5; i.e., for all workers, for recently absent workers, and for recently present workers. For comparison, we also show the estimated linear relationships (from Table 5), and to illustrate the origins of the linear slope estimates, we present the new dummy-based estimates with circles representing the numbers of GP-months that lie behind each estimate, as well as with 95 % confidence intervals. The estimates in Figure 4 suggest indeed that the association between worker utility and GP leniency is non-linear, and that, for most workers, the relationship is relatively flat in the central part of the GP leniency distribution. However, for all worker groups, there appears to be a negative utility associated with having a very strict GP. For recently/currently absent workers there also appears to be a considerable utility gain associated with having a very lenient GP; see panel (b). It is notable that for the latter group – which clearly constitutes the worker group for which GP leniency is likely to have some immediate consequences – the relationship appears to be highly monotone, and also not very far from the linearity assumption.

Our reading of these exercises is that our finding that workers put considerable emphasis on a GPs leniency reputation when choosing a new family doctor is robust with respect to modeling assumptions. The relationship is strongest in the tails – and particularly in the lower tail – of the GP leniency distribution.

4.4 The alternative leniency index

As explained in Section 3, our GP leniency indicator also picks up fluctuations in the health conditions of the GPs' customers. Since we have designed our analysis such that the workers choosing a new GP have never contributed to the calculation of the leniency indicators for the GPs included in their choice sets, this does not by itself compromise the causal interpretation of our findings in this section. However, if the mechanism that triggers a new GP choice is systematically driven by health shocks that also have implications for the actual choice of GP, this may bias our results. To address this concern, we re-estimate our choice model using the alternative leniency indicator, based on *predicted* rather than *actual* customer lists; see Section 3.2. This is done for the complete set of choice situations, as well as for the subpopulations distinguished by past absence behavior. The results are provided in Table 7, which is directly comparable to Table 5 above. The message coming out of this exercise is that the estimated impacts of GP leniency are reduced considerably, as we would expect from a much noisier indicator, but that they still are statistically significant and display the same pattern of larger impacts for workers with more absence. The main findings reported in Section 4.2 thus appear to be robust also with respect to the simultaneity problem caused by possibly systematic relationships between health shocks and GP selection.

Table 7. Workers' choice of new GP. Estimation results base on alternative leniency index (standard errors in parentheses)

	(1) All	(2) Absent t-1	(3) Present t-1	(4) Present up to GP choice	(5) Always present
GP leniency	0.072*** (0.004)	0.139*** (0.010)	0.059*** (0.005)	0.045*** (0.008)	0.036*** (0.011)
Within walking distance	0.577*** (0.006)	0.519*** (0.016)	0.583*** (0.007)	0.656*** (0.012)	0.655*** (0.017)
Estimated driving time if not walking distance (minutes)	-0.107*** (0.000)	-0.098*** (0.001)	-0.109*** (0.001)	-0.113*** (0.001)	-0.119*** (0.001)
Emergency service provided	0.013*** (0.004)	0.021** (0.010)	0.012*** (0.004)	-0.028*** (0.007)	-0.030*** (0.010)
Same sex GP	0.661*** (0.004)	0.546*** (0.010)	0.685*** (0.004)	0.729*** (0.007)	0.718*** (0.011)
Female GP	-0.095*** (0.004)	-0.225*** (0.010)	-0.068*** (0.005)	-0.028*** (0.008)	-0.046*** (0.012)
GP age	-0.020*** (0.000)	-0.025*** (0.001)	-0.019*** (0.000)	-0.013*** (0.000)	-0.011*** (0.001)
GP younger than patient	0.009 (0.008)	-0.060*** (0.019)	0.023*** (0.009)	0.049*** (0.017)	0.082*** (0.023)
GP older than patient	-0.130*** (0.006)	-0.089*** (0.016)	-0.139*** (0.007)	-0.150*** (0.012)	-0.169*** (0.017)
Free capacity at start of month	0.989*** (0.004)	1.033*** (0.010)	0.981*** (0.004)	1.176*** (0.007)	1.187*** (0.011)
N #choices×size of choice set	28,178,517	4,597,138	23,581,379	9,118,840	4,370,809

Note: GP leniency is defined a in Eq. (5), but based on predicted instead of actual GP-worker affiliations, as described in Section 3.2. Observation numbers deviate slightly from Table 5 for the reason that some GPs were not attributed a sufficient number of employed customers by our prediction model. Walking distance is defined as less than 1,500 meters (0.9 miles). Estimated driving times are computed from "open street map". ***(**)(*) indicates statistical significance at the 1(5

4.5 Termination of existing GP relationships

While the analysis in this section has been conditional on the termination of an existing customer-GP relationship, we may suspect that these decisions are also influenced by assessments of the current GP's leniency in comparison to feasible alternatives. This is more difficult to model empirically, however, as new choices can be made at any point in time and as the sets of alternative GPs and their leniency behaviors are constantly changing. Decision to leave an existing, and previously selected GP, may also have been triggered by conflicts or episodes causing dissatisfaction that are unobservable to us. In particular, it is probable that a worker with a very high demand for sick leave certificates is met with less leniency over time, as the GP may start suspecting that the patient is exploiting his/her leniency.

To examine the extent to which changes in a GP's leniency behavior affects the decision to leave an existing GP we set up a simple linear probability model for all worker-months with an existing family doctor affiliation (approximately 134 million observations), where the outcome is a dummy variable equal to one if a switch occurs during the month. We clearly cannot use as explanatory variable either of the leniency indicators computed in Section 3, as these have been affected by the workers' own absence behavior in the past. Instead, we focus on the more recent changes in the GPs absence certification for his/her other customers. Specifically, we apply a leave-one-out mean for the monthly number of absence days prescribed for *all the other* workers affiliated to the same GP over the past three months (strictly prior to the month in question). In addition, we include the corresponding number of absence days prescribed for the worker himself/herself over the same period. While the former of these variables is included to capture changes in the GP's leniency behavior more generally, the latter is included to encapsulate changes in patient-specific leniency. To ensure that our findings are not driven by cross-sectional variations in absence and GP shifting behavior and/or by correlation between individual absence and GP shifting tendencies, we include worker fixed effects in this model. We also check robustness by adding GP fixed effects, in essence focusing exclusively on within-GP variations in absence certification. Finally, in order to incorporate the idea that the existing GP's practice style may be evaluated relative to the most likely alternatives, we estimate a model where we use the leave-one-out mean for own GP normalized by the corresponding mean for all workers in the local area regardless of GP.

Table 8. Workers' decisions to leave a current GP. Linear probability estimates multiplied by 100 (standard errors in parentheses)

	(1)	(2)	(3)
Own absence (days per month)	0.0070***	0.0086***	0.0084***
	(0.0002)	(0.0002)	(0.0001)
Absence prescribed for other workers			
Absolute level (av. days per month)	-0.2163***	-0.2171***	
	(0.0201)	(0.0002)	
Relative to others in the local area			-0.1685***
			(0.0037)
Worker fixed effects	Yes	Yes	Yes
GP fixed effects	No	Yes	Yes
N (# monthly observations)	134,315,536	134,315,536	134,315,536

Note: The outcome variable is equal to unity in 0.6 % of the cases. The explanatory variables have the following means (standard deviations): Own absence: 1.72 (6.05), Absence prescribed for other workers: 1.79 (0.56), Absence relative to others in the local area: 1.02 (0.29). The reason why the number of monthly observations is lower than the 158 mill. observations used to compute the leniency index in Section 3 is that we have dropped from the analysis physicians who are not genuine family doctors; see footnote 6. Standard errors are two-ways clustered on worker and GP. ***(**)(*) indicates statistical significance at the 1(5)(10) % levels, respectively

The results are provided in Table 8. They confirm that a more lenient GP behavior (as reflected in the leave-one-out mean of prescribed absence days) reduces the probability of leaving the GP. They also confirm that high own absence in the past raises the probability of leaving the current GP, possibly because the GP becomes more skeptical with respect to the demand for additional sick leave prescription. The results are robust with respect to model specification.

The finding that higher own absence in the recent past also imply a higher probability of shifting to a new family doctor suggests that workers who have a lot of absence tend to switch GP more often than workers who are rarely absent from work, given that they do not move to another location. As shown in Figure 5, this is indeed the case. Although the differences are small in quantitative terms, there is a monotonous pattern of higher switching propensity the higher is the overall number of absence days, and the most absence-prone workers switch approximately 25 % more often than the workers without recorded sick leave.

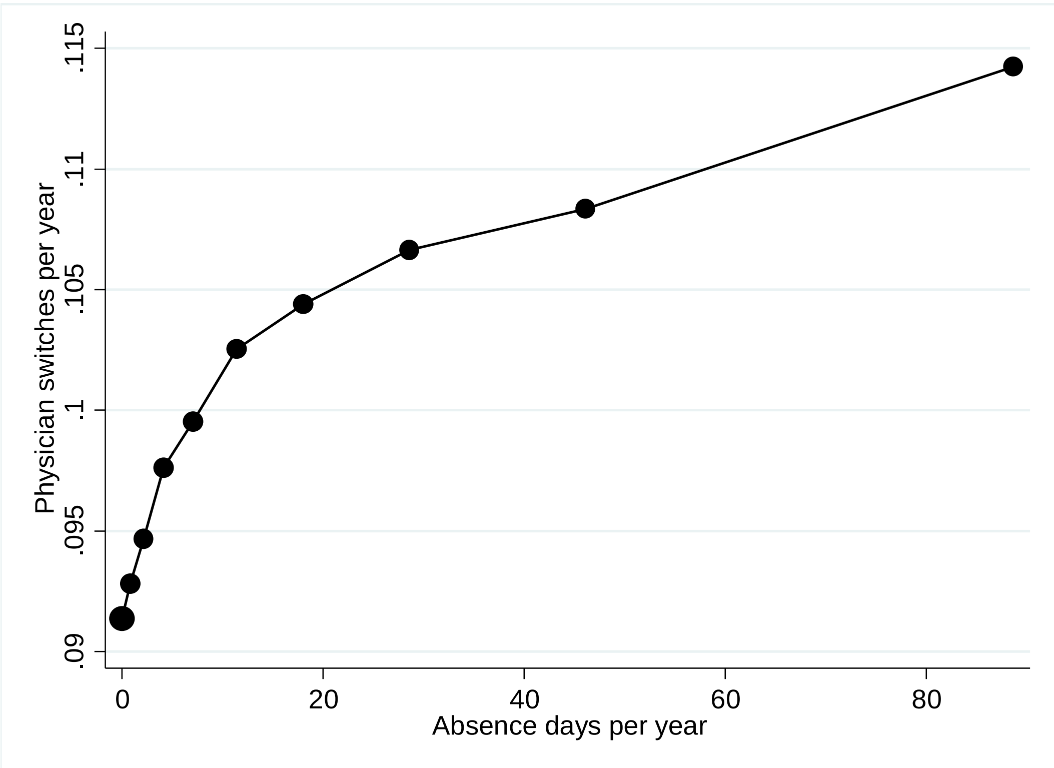

Figure 5. Yearly number of absence days and GP switches.
Note: The graph describes the population of 1.2 million workers who are employed for at least five years in total and remain resident in the same local area over the data period. The population is divided into 10 groups, based on their total number of absence days per year. Those with zero absence constitute one group (190,000 workers), whereas the other nine groups are equally sized (115,000 workers in each group).

5 The GPs' choice of leniency

Having established that workers do take the GPs' leniency reputation into account when selecting their family doctor, the next question is whether or not the physicians respond to the demand for lenient gatekeeping practices in order to attract customers? If they do, we would expect leniency to increase in response to patient shortage or increased GP market competition, at least for GPs whose earnings depend on the number of customers. We would also expect GPs running their own business to be more lenient than GPs on a fixed wage contract, ceteris paribus. In this section, we first present a descriptive overview of the observed statistical associations between estimated leniency and market conditions as reflected in the type of remuneration and the degree of local patient competition. We then turn a statistical analysis aimed at identifying and estimating the causal impacts of these conditions on GPs' gatekeeping leniency.

5.1 Fixed or variable wages

As described in Section 2, whereas most of the family doctors in Norway run their own business, and hence have earnings that depend on the number of customers and consultations/treatments, approximately 4.5 % of the family doctors are on fixed-wage contracts. Figure 6 shows density plots for the estimated monthly leniency indicator from Section 3.1 separately for fixed-wage physicians and physicians running their own business (hereafter referred to as variable-wage physicians). Although there is a considerable variation in estimated gatekeeping practices within each of these groups, it seems evident that variable-wage physicians tend to be a bit more lenient. The difference in average leniency between the two physician groups is 0.12, suggesting that variable-wage physicians certify 0.12 (7 %) more absence days per month than fixed-wage physicians do, ceteris paribus.

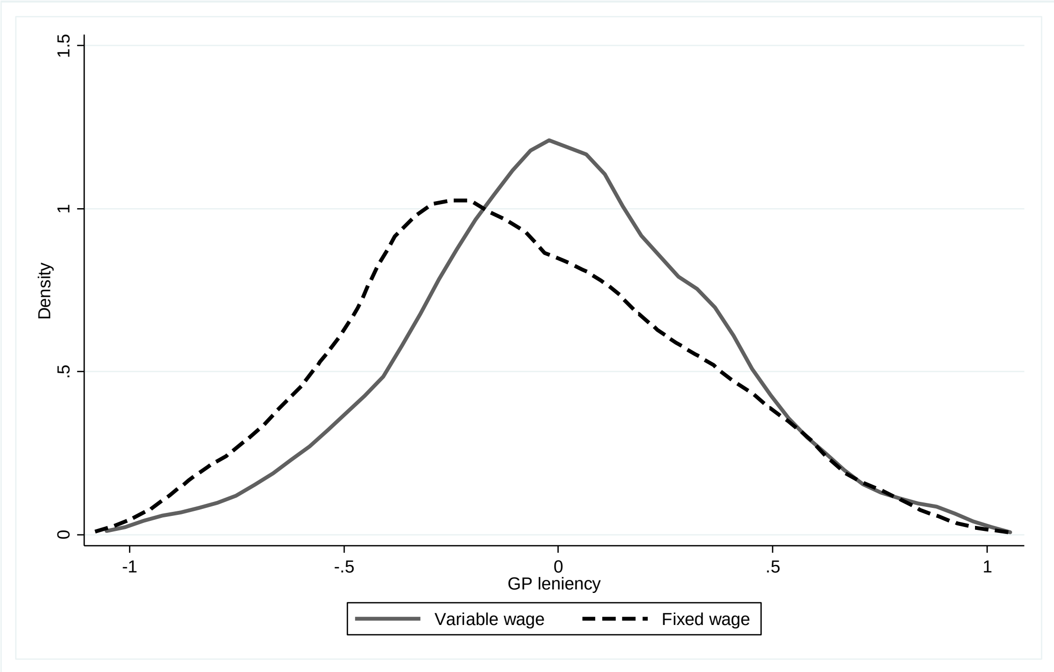

Figure 6. . The distribution of time-averaged leniency indicators for the physicians in Norway by type of remuneration
Note: The density plots are based on the weighted distribution of the time-averaged leniency indicators computed in Section 3.1 (with number of employed customers used as weights). Physicians who shift between variable and fixed-wage contracts during the data period are in this graph treated as two different physicians.

5.2 The competitive environment

In order to examine the physicians' leniency-responsiveness with respect to their competitive situation, we take a closer look at the physician markets identified in the previous section, and check whether the average leniency indicators among variable-wage physicians correlate systematically with the degree of local competition. In this descriptive exercise, we include all local areas with at least 200 inhabitants divided between at least 10 physicians who each have at least 10 employed customers from the local area. This gives us approximately 105,000 monthly local physician-market observations. Since it is not obvious how the degree of local competition between physicians should be measured, we compute for each local area and for each calendar month, a number of alternative competition indicators; i.e.:

(a) The vacancy rate, defined as the number of vacant (unfilled) patient slots in the local area divided by the total number of registered (desired) patient slots. This is perhaps the most natural local competition indicator, as it summarizes the overall excess supply of patient slots. Note that in some cases, the vacancy rate is negative, as it is possible for physicians to be temporarily overbooked.

(b) The number of vacant patient slots in the local area. This measure also summarizes overall excess supply, but in absolute numbers instead of relative to total supply as in (a).

(c) The number of GPs in the local area with at least one vacant patient slot (available physicians). This is the best indicator for the actual choice opportunities available to workers searching for a new GP.

(d) The number of physicians operating in the local area divided by the number of inhabitants. This is the standard measure of physician density.

(e) Patient turnover, defined as the fraction of patients in the local area who shifted family doctor last month. This measure captures the current tendency for patients in the local area to shift family doctor.

(f) The fraction of physicians with fixed-wage contracts. In contrast to the other competition indicators, a higher fraction of fixed-wage physicians indicate *less* competition, both because fixed-wage contracts tend to be offered precisely *because* the GP coverage is considered too low and because fixed-wage physicians do not have the same incentives to compete for patients.

Figure 7 illustrates in separate panels (a)-(f) how each of these competition indicators correlates with average estimated leniency among the local areas' variable-wage physicians. For ease of exposition, we have divided the local market observations into percentiles (with approximately 1,000 market observations behind each data point). To illustrate the considerable size-differences between the local markets, we have let the sizes of the dots vary proportionally to the total number of inhabitants in the local markets included in each percentile. Finally, we have added linear bivariate regression lines, based on the size-weighted percentiles.

It is clear from Figure 7 that, regardless of the chosen competition indicator, we consistently uncover a conspicuous and robust positive relationship between the degree of local competition and average estimated physician leniency. It appears that the gradient is steeper at lower levels of competition, suggesting that while there is a relatively strong effect on leniency of introducing a bit competition in a local market, the marginal effect flattens out as the competition becomes more intense. However, the relatively few data points that we have at high competition levels makes it difficult to draw any clear conclusions regarding the functional form relationship between local competition and GP leniency.

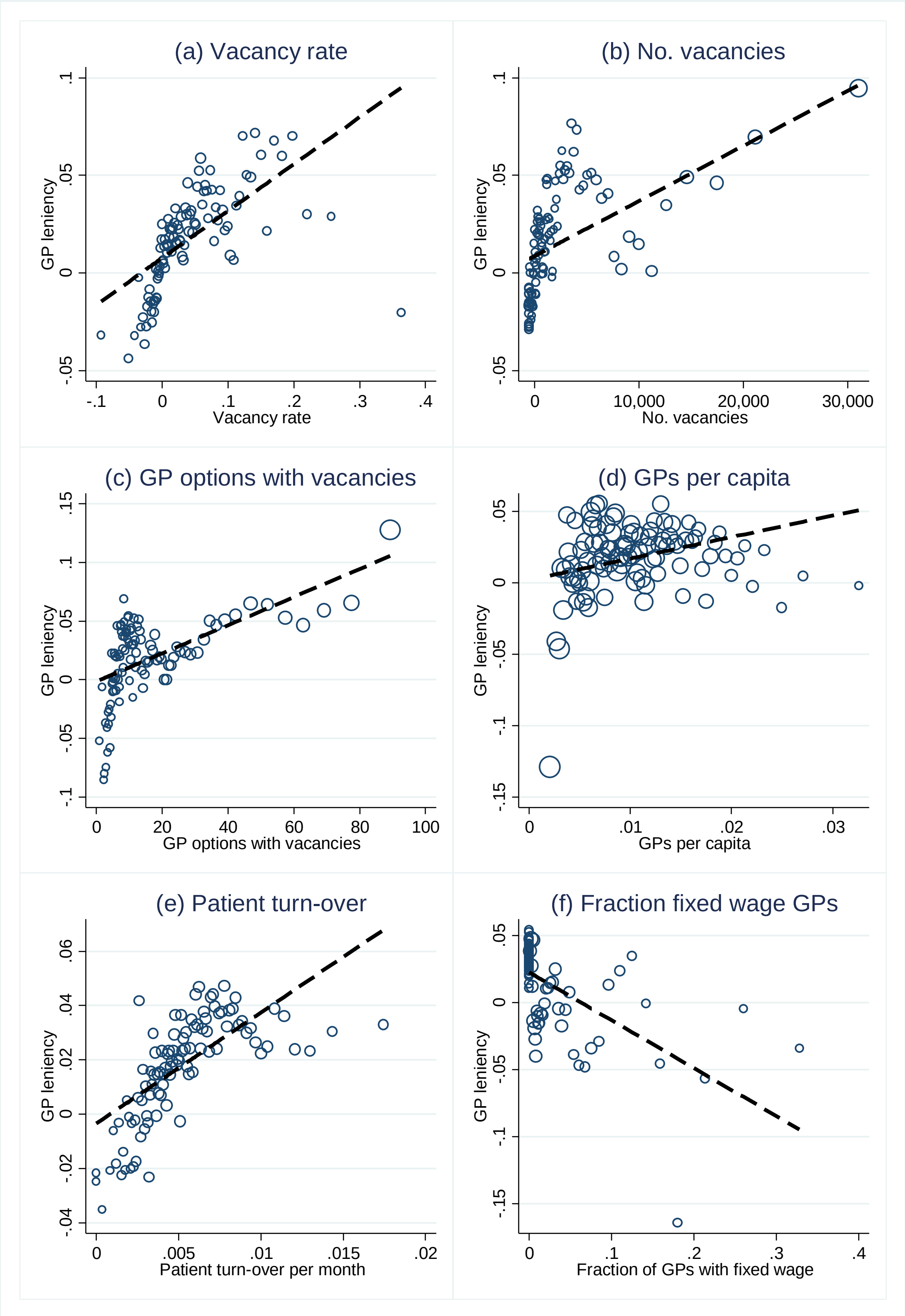

Figure 7. Average leniency indicator among variable-wage physicians in local physician markets. By degree of competition.

Note: Each data point represents a percentile in the respective distributions of competition indicators. The sizes of the dots are proportional to the number of inhabitants in the local markets included in each percentile.

5.3 Causal analysis

If the largely cross-sectional correlation patterns in Figure 7 have a causal interpretation, we would expect to find them also in a longitudinal setting; i.e., we would expect each physician's gatekeeper leniency to change in response to changes in the local competitive environment and own patient shortage. To investigate this further, we examine in more detail the individual physicians' responses to changes in their competitive situation. The data we use for this purpose is described in Table 9. They consist of panel data for 5,340 family physicians, on average observed for 71 months during our estimation period. On average, a physician is 56 customers short; i.e., he/she has 56 fewer persons on the customer-list than desired, corresponding to an average vacancy rate of 5.3 %. However, the vacancies are far from equally distributed, and in as much as 55 % of the physician-month observations, the list is full or overbooked. And the median GP has five patients more than actually desired.

Table 9. GPs in Norway January 2002 - December 2010

Total number of physicians	5,340				
Average number of months represented in the dataset	71.05				
Descriptive statistics by physician month	Mean	Standard deviation	10^{th} percentlie	Median	90^{th} percentile
Number of desired patients	1,410.9	399.5	950	1380	1980
Number actual patients	1,355.0	386.4	888	1339	1853
Number of vacancies	56.1	198.4	-67	-5	267
Fraction of GPs with full lists (no vacancies at start of month)	0.554				
Vacancy rate (vacancies/number of desired patients)	0.053	0.175	-0.050	-0.004	0.235
Fraction on fixed-wage contract	0.476				
Local market competition (local areas)					
Vacancy rate	0.058	0.066	-0.005	0.040	0.150
No. vacancies	8,593.4	14,946.9	-210.8	2237.9	29,087.7
No. GP options with vacancies	21.13	24.16	3.21	10.79	60.09
GPs per capita	0.0007	0.0001	0.0006	0.0007	0.0008
Patient turn-over per month	0.006	0.004	0.003	0.005	0.009
Fraction of GPs with fixed wage	0.045	0.084	0.0003	0.022	0.105

Note: All numbers are weighted by the number of workers behind each GP/month measure of leniency.

Note: A negative vacancy number implies that the GP is overbooked, in the sense that he/she currently has more patients than desired.

To examine the determinants of gatekeeper leniency, we set up and estimate statistical models with the following structure:

$$\hat{\mu}_{jt} = \pi_j + \mathbf{z}_{\mathbf{jt-1}}\boldsymbol{\sigma} + \varepsilon_{jt}, \tag{7}$$

where $\hat{\mu}_{jt}$ is the monthly leniency index defined in Equation (2) and $\mathbf{z}_{\mathbf{jt}}$ is a vector of time-varying variables that potentially influence the physicians' practice style. To ensure that the presumed endogenous responses take place strictly after the explanatory impulses, we enter all the explanatory variables with a one month time lag. While we focus on the impacts of the type of wage contract and local market conditions, all estimated models also include dummy variables for calendar time (on a monthly basis) and the physician's age (on a yearly basis). In a baseline specification, we represent market conditions as the average vacancy rate among the physicians in the local market (including all physicians who have at least 10 patients from the local area, irrespective of where the physicians themselves are located). The variables are entered as deviations from their respective (global) means.

Since we use estimated coefficients (from Equations (1) and (2)) as the dependent variable in (7), there is considerable statistical noise in this variable arising from this first-step estimation. This noise also varies with the precision by which the various physician-by-month leniency indicators have been estimated, inducing heteroskedasticity in the residuals. As described in the previous section, we have performed a nonparametric bootstrap for the calculation of leniency indicators, with 104 trials. Based on this bootstrap, we estimate the variance applying for each physician-by-month leniency indicator, and then estimate Equation (7) by weighted least squares (WLS) with the inverse of these estimated variances used as weights. Finally, in order to obtain correct standard errors for the parameters of interest, we perform an additional nonparametric bootstrap for the estimation of (7); i.e., for each estimated vector of $\hat{\mu}_{jt}$, we randomly draw (with replacement) a new sample of estimated leniency indicators, and re-estimate Equation (7). We then use the standard deviations computed for each of the respective coefficient estimates as estimates of their standard errors.

5.4 Main results

The results are presented in Table 10. As a point of departure, columns (1) and (2) first show results from a simple weighted least squares (WLS) regression and a "between-estimator" (using the cross section of physician averages only), where we include the fixed-wage indicator and the local vacancy rate only. Columns (3)-(5) show the results from the fixed effects model, with the three key variables entered in a stepwise fashion. Column (3) shows the estimated impact of having a fixed rather than a variable wage, with no controls for market situation or own vacancies. The point estimate of -0.071 implies that when a physician switches from a variable to a fixed-wage regime, the expected number of certified absence days among his/her customers drop by -0.071 per month – corresponding to approximately 4.0 % of the average number of prescribed days; conf. Table 1. This is considerably smaller than suggested by the pure cross-sectional regression in column (2) and the OLS model in column (1), yet still significant from a substantive point of view. The foundation for identification is weak, however, as the fixed-wage coefficient estimated in a model with physician fixed effects is based on the relatively few physicians who change between a fixed wage and a variable wage status (47 physicians in our data). Column (4) presents the results from a model where we have included the local vacancy rate (for variable-wage physicians only). The point estimate of 0.23 implies that if the local vacancy rate increases by, say, 0.1 (10 percentage points), the average monthly number of physician-certified sick pay days prescribed by variable-wage physicians in that local area increases by approximately 0.023 days.

Table 10. Estimation results. Weighted Least Squares.
Dependent variable: Estimated GP leniency (standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)
	Between and within GPs	Between GPs	Within GPs	Within GPs	Within GPs	Within GPs
Fixed wage (dummy)	-0.114*** (0.012)	-0.156*** (0.021)	-0.071* (0.043)	-0.061 (0.043)	-0.058 (0.043)	-0.058 (0.044)
Variable wage (dummy) interacted with:						
Own vacancy rate					0.063*** (0.015)	0.062*** (0.015)
Local market vacancy rate	0.118** (0.055)	0.138 (0.097)		0.226*** (0.037)	0.184*** (0.040)	0.184*** (0.040)
Fixed wage (dummy) interacted with						
Own vacancy rate						0.015 (0.063)
Local market vacancy rate						-0.079 (0.114)
N (physician-months)	379,425	379,425	379,425	379,425	379,425	379,425

Note: Local market vacancy rate is computed in a "leave-out"-fashion; i.e., without including data from own practice. All models include calendar month and years of age dummy variables. "Fixed wage" is a dummy variable indicating that the GP receives a fixed wage from the municipality. "Variable wage" is a dummy variable indicating that the wage is not fixed. Columns (3) - (6) are based on models with physician fixed effects included. Standard errors are based on nonparametric bootstrap in two steps, each consisting of 104 trials. The first step consists of resampling individual workers and re-estimating the leniency indicator (Equations (1) and (2)). The second step consists of resampling physician month observations and re-estimating the model for the GP's choice of leniency (Equation (7)).***(**)(*) indicates statistical significance at the 1(5)(10) % levels, respectively.

Column (5) shows our preferred results where we also add in the physician's own vacancy rate. The estimated coefficient for this variable is considerably smaller than for the overall vacancy rate in the local market, yet highly statistically significant. Finally, column (6) adds in the local market and own vacancy rate for fixed-wage physicians. Since these physicians do not have any personal economic motives for adjusting their gatekeeping practices in response to market conditions, this model specification may be viewed as a sort of placebo test. Given the low number of fixed-wage physicians, the standard errors are considerably larger for these coefficients; hence statistical power probably becomes too weak to draw firm conclusions. However, the two estimated coefficients are either approximately zero or "wrongly" signed, and none of them are statistically significant.

5.5 Robustness

How robust are these results? Table 11 reports estimation results for a range of alternative fixed effects models. For ease of comparison, column (1) repeats the main results from column (5) in Table 10. Columns (2)-(6) report results based on alternative formulations of the competitive pressure variables. In column (2), we have substituted the absolute *number* of own and local vacancies for the corresponding *rates*. In column (3) we have kept the own vacancy rate, but used the number of GPs with vacant patient slots as the indicator for local competition. In column (4), we have instead used the overall number of GPs per capita in the local area as the indicator for local competition. In column (5) we have used the fraction of persons who has recently shifted family doctor. And in column (6), we have used the fraction of GPs in the local area with variable wage. All these models essentially give the same answers: More competition implies more lenient gatekeepers. Finally, column (6) reports how the estimates in the baseline model change when we add in a lagged dependent variable in Equation (7).[16] This reduces the estimated coefficients slightly. The "steady state" effects (the coefficient estimate divided by one minus the estimated autoregressive parameter) remain approximately unchanged, however

The message coming out of these exercises is that our main finding – that gatekeeping practices are indeed affected by market conditions and the physicians own economic incentives – is highly robust with respect to model specification.

[16] Note that the transformed equation used to estimate the fixed effects model contains a residual that incorporates a covariate-adjusted individual mean (over all months), and is thus not completely exogenous with respect to the lagged dependent variable (see, e.g., Cameron and Trivedi (2005, p. 764)). Approximate consistency requires that the average residual is small relative to each period's residual, which again requires that the number of periods is large. We consider this satisfied in our case.

Table 11. Estimation results from alternative model specifications with physician fixed effects.
Dependent variable: Estimated GP leniency (standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fixed wage (dummy)	-0.058 (0.043)	-0.078* (0.043)	-0.085** (0.043)	-0.054 (0.044)	-0.065 (0.043)	-0.076* (0.044)	-0.046 (0.040)
Variable wage (dummy) interacted with:							
Own vacancy rate	0.062*** (0.015)		0.069*** (0.014)	0.072*** (0.014)	0.076*** (0.014)	0.076*** (0.014)	0.045*** (0.014)
Local market vacancy rate	0.184*** (0.039)						0.137*** (0.038)
Number of own vacancies (divided by 100)		0.004*** (0.001)					
Number of local market vacancies (divided by 5000)		0.008*** (0.001)					
Number of GPs with vacancies (divided by 100)			0.125*** (0.028)				
Number of GPs divided by number of inhabitants				0.117*** (0.034)			
Fraction of customers with recent GP shift					0.596 (0.409)		
Fraction of GPs with variable wage						0.125** (0.060)	
Lagged dependent variable							0.258*** (0.002)
N (physician-months)	379,425	379,425	379,425	379,425	379,425	379,425	379,425

Note: Where appropriate, local market characteristics are computed in a "leave-out"-fashion, i.e., without including data from own practice. All models include calendar month and years of age dummy variables. "Fixed wage" is a dummy variable indicating that the GP receives a fixed wage from the municipality. "Variable wage" is a dummy variable indicating that the wage is not fixed. Physician fixed effects included in all models. Standard errors are based on nonparametric bootstrap; se note to Table 10.
***(**)(*) indicates statistical significance at the 1(5)(10) % levels, respectively.

6 The association between leniency and customer numbers

Somewhat simplified, we have found that: i) more lenient gatekeeping gives the GP more customers, and ii) more customers make the GP less lenient. What are the overall implications of these two mechanisms for the association between GP leniency and the allocation of customers? Will the lenient physicians tend to have more or less customers than the strict physicians? One way to address this question is to examine the statistical association between the predicted gatekeeper leniency and the GPs' numbers of employed customers, using all variable-wage GP-months as the units of observation.

Let n_{jt} denote the number of employed customers held by family doctor *j* in a month *t,* and recall that $\hat{\mu}_{jt}$ denotes the GP's estimated leniency in the same month. Based on all the GP-month observations in our dataset, we find that the correlation between these two variables is negative $(corr(\hat{\mu}_{jt} n_{jt}) = -0.08)$. This implies that lenient physician behavior is empirically associated with fewer rather than more customers. However, in order to give this association a behavioral interpretation, we need to disentangle the within-market and across-market contributions to the correlation pattern. Across markets, we would indeed expect the association to be *negative,* as a high (low) total number of customers per GP in a market implies a low (high) degree of competition, and thus unambiguously yield less (more) lenient GP behavior (confer Figure 7). To facilitate an appropriate decomposition into within-market and across-market contributions, we assume that all variable-wage GPs located within a given local area in a given month constitute a unique market.[17] We then obtain a set of non-overlapping markets and can use the law of total covariance to disentangle the within-market and the across-market contributions. Let Z_{jt} indicate the market in which GP *j* operates in moth *t*. We then have that

$$\underbrace{\operatorname{cov}(\hat{\mu}_{jt}, n_{jt})}_{=-7.49} = \underbrace{E\left[\operatorname{cov}(\hat{\mu}_{jt}, n_{jt} \mid Z_{jt})\right]}_{=-5.51} + \underbrace{\operatorname{cov}\left[E(\hat{\mu}_{jt} \mid Z_{jt}), E(n_{jt} \mid Z_{jt})\right]}_{=-1.98}, \tag{8}$$

where the first expression on the right-hand-side reflects the within-market covariance between leniency and the number of customers, and the second expression reflects the across-market covariance. The numbers entered below the expressions show the corresponding estimates computed from our data. Hence, what we find is that the negative correlation between GP leniency

[17] This is obviously not correct, as we have shown in this paper that there is a considerable overlap between markets, such that many workers have GPs outside their own local area. For the purpose at hand, however, we consider it to be an acceptable approximation. Note also that we have identified each physician's locality as the local area in which the majority of his/her customers live.

and customer numbers arises both *within* and *across* local markets, reflecting that the GP-responses dominate the customers' choice behavior.

Another way of illustrating the consequences of customer choices and leniency adjustments is to compare the unweigthed average GP leniency (taken over GP-by-month-observations) with the corresponding average weighted by the number of employed customers each month. The result of this exercise is presented in Table 12. The first row of this table reports the average of the prescribed absence days taken over all GP-months (without weighting by number of customers). The second row reports the corresponding average weighted by the number of employed customers. A comparison of the two numbers confirms that the average *chosen* physician is slightly *less* lenient than the average physician. The mechanism behind this result is further illustrated by looking at the predicted GP-contributions to absence prescription conditional on "popularity". A conspicuous finding is that the GPs with full customer lists are much less lenient than GPs who have fewer customers than desired. Taken literally, the estimates imply that had everyone been assigned the most popular physicians – those with full customer lists more than 95 % of the time – absenteeism would have been reduced by as much as 5.5 %. The obvious explanation for this is that the most lenient physicians are lenient precisely because they have difficulties recruiting a sufficient number of customers. Hence, somewhat simplified, we could say that lenient gatekeeper practice appears to be a tool used to compensate for other characteristics that entail a low probability of being selected.[18]

Table 12. Average prescribed absence days implied by estimated leniency. By physicians' popularity

	Implied average
Unweighted average of absence days by GP (N=379,425)	1.83 days
… weighted by number of customers each month	1.82 days (-0.74 %)
… conditional on full customer list (N=198,098)	1.82 days (-0.73 %)
… conditional on vacant customer slots (N=181,327)	1.85 days (+0.88 %)
… conditional on full lists ≤ 75 % of the time (N=246,105)	1.85 days (+1.03 %)
… conditional on full lists > 75 % of the time (N=133,320)	1.80 days (-1.91 %)
… conditional on full lists > 95 % of the time (N=48,445)	1.73 days (-5.52 %)

Note: All reported percentage changes are computed from the respective differences in estimated GP leniency only, and measured relative to the average number of absence days per month reported in the first row (1.83). This number deviates from the average number reported in Table 1 (1.79) for two reasons: i) While the average reported in Table 1 is taken over workers, the average reported here is taken over physicians (unweighted by number of workers), and ii) while Table 1 includes absence prescribed by all physicians, the present table includes absence prescribed by genuine family physicians (where a specific person can be identified); see footnote 5.

[18] Previous evidence from the Norwegian GP system has indicated that physicians with fewer customers than desired also are deselected 50 % more frequently than physicians with full lists are (Iversen and Lurås, 2011).

7 Concluding remarks

Does the assignment of GPs to the twin roles of running competitive businesses and guarding the public purse imply that we in practice have “let the fox mind the henhouse”? We have shown in the previous sections that workers in the process of choosing a new family doctor prefer lenient over strict gatekeepers, ceteris paribus. We have also shown that some physicians take this customer-preference for leniency into account when determining their own gatekeeping practice style, such that they become more lenient when the competition for patients become fiercer. But how important are these mechanisms? What are the financial implications for the insurer (the taxpayer in our case)?

We can shed some light on the issue by using the physicians on fixed-wage contracts as a point of reference. As shown in Section 3, the average difference in estimated leniency between variable and fixed-wage physicians is 0.12, which corresponds quite closely to the WLS estimate in column (1) of Table 10. However, based on the fixed effect (within GP) model, we have estimated that approximately half of this difference has a causal interpretation. Taken literally, this suggests that if we – as a thought experiment – gave all the GPs in Norway a fixed-wage contract instead of letting them run their own businesses, we would expect the absence level in Norway to decline by approximately 0.06 days per month, or roughly 3-4 %. This is a considerable, though not huge effect.

Why do we not see even larger consequences of letting the GPs be gatekeepers for their own customers? After all, the responses identified in this paper, both on workers’ GP choices and on GP behavior, appear both statistically and substantively significant. One important explanation is that the competition between physicians in Norway is weak. In large parts of the country, GPs are persistently in short supply, and as shown above, more than half of the GPs’ patient lists are fully booked (or overbooked) while the average vacancy rate is only 5 % (Table 9). And even though workers care about physician leniency, they also care about other physician characteristics. Hence, many GPs probably do not have to worry at all about insufficient demand for their services. And the most popular GPs need not be particularly lenient in order to attract the desired (or the maximum) number of customers. In fact, family doctors in Norway on average work 20 % more hours per week than the 37.5 hours defined as full time; and in a survey conducted in 2004, a third of the family doctors reported that they were subjected to an “unacceptable” level of work pressure (Aasland and Rosta, 2011).

The estimated effects are indeed larger if we focus on markets with above average competition. Combining our preferred estimates (Table 10, column 5) with the vacancy rates prevailing in the local markets with most intense physician competition (with vacancy rates around 30 %), we find that substituting fixed wage for all variable-wage physician would reduce the expected absence level by approximately 7 %.

Although the findings reported in this paper hardly suggest that strict gatekeepers are necessarily driven out of business, they illustrate the potential difficulty associated with being an impartial gatekeeper for claims made by own customers. And this role conflict can extend far beyond the certification of sick pay. In most countries, family doctors play important gatekeeper roles in choices of (costly) medical treatments, in decisions regarding referrals to specialists, and in the evaluation of disability insurance claims. The fiscal consequences of this system are likely to depend on the degree of competition between GPs; the more physicians need to worry about attracting enough customers, the more costly it is to use them as the gatekeepers of social or private insurance.

References

Aasland, O. G. and Rosta, J. (2011) The working hours of general practitioners 2000 – 2008. *Tidsskrift for Den norske legeforening,* Vol. 131, 1076-1079.

Bonato, L. and Lusinyan, L. (2007) Work Absence in Europe. IMF Staff Papers, 54, 475-538.

Cameron, A. C. and Trivedi, P. K. (2005) *Microeconometrics – Methods and Applications*. Cambridge University Press.

Carlsen, B. and Norheim, O. F. (2005) “Saying No Is No Easy Matter” A Qualitative Studt of Competing Concerns in Rationing Decisions in General Practice. *BMC Health Services Research,* Vol. 5, No. 70.

Carlsen, B., Nyborg, K. (2009) The Gate Is Open: Primary Care Physicians as Social Security Gatekeepers. Memorandum 07/2009. Oslo: Department of Economics, University of Oslo.

Chandra, A., Cutler, D. and Song, Z. (2012) Who Ordered That? The Economics of Treatment Choices in Medical Care. In *Handbook of Health Economics,* Vol. 2, eds. McGuire, T. G., Pauly, M. V., and Barros. P. P., 397-432. Elsevier.

Dusheiko, M., Gravelle, H., Jacobs, R., Smith, P. (2006) The Effect of Financial Incentives on Gatekeeping Doctors: Evidence from a Natural Experiment. *Journal of Health Economics* Vol. 25, 449-478.

Englund, L. and Svärdsudd, K. (2000) Sick-listing habits among general practitioners in a Swedish county. *Scandinavian Journal of Primary Health Care,* Vol. 18, 81-86.

Gaure, S. (2013) OLS with Multiple High Dimensional Category Variables. *Computational Statistics & Data Analysis*, Vol. 66, 8–18.

Gaure, S. (2014) Correlation Bias Correction in Two-Way Fixed-Effects Linear Regression. *Stat*, Vol. 3, No 1, 379–390.

Godager, G., Iversen, T. and Ma, C.A. (2015) Competition, Gatekeeping, and Health Care Access. *Journal of Health Economics*, Vol. 39, 159–170.

Gowrisankaran, G., and Town, R. J. (1999) Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics,* Vol. 18, 747-767.

Grytten, J. and Sørensen, R. (2003) Practice Variation and Physician-Specific Effects. *Journal of Health Economics,* Vol. 22, 403-418.

Inspektionen för socialförsäkringen (2014) Vårdvalets effekter på sjukskrivningarna (The effect of introducing a health care choice system on the use of sickness insurance). Rapport 2014:17.

Iversen, T. (2004) The Effects of Patient Shortage on General Practitioners' Future Income and List of Patients. *Journal of Health Economics,* Vol. 23, 673-694.

Iversen, T. and Lurås, H. (2000) Economic Motives and Professional Norms: The Case of General Medical Practice. *Journal of Economic Behavior & Organization*, Vol. 43, 447-470.

Iversen, T. and Lurås, H. (2011) Patient Switching in General Practice. *Journal of Health Economics,* Vol. 30, 894-903.

Iversen, T., Ma, C.A. (2011) Market Conditions and General Practitioners' Referrals. *International Journal of Health Care Finance and Economics*, Vol. 11, 245–265.

Markussen, S., Mykletun, A. and Røed, K. (2012) The Case for Presenteeism - Evidence from Norway's Sickness Insurance Program. *Journal of Public Economics*, Vol. 96, 959-972.

Markussen, S., Røed, K., Røgeberg, O.J., Gaure, S. (2011) The anatomy of absenteeism. *Journal of Health Economics,* Vol. 30, 277-292.

Markussen, S, Røed, K. and Røgeberg, O. (2013) The Changing of the Guards - Can Family Doctors Contain Worker Absenteeism? *Journal of Health Economics*, Vol. 32, No. 6, 1230–1239.

McFadden, D. L. (1974) Conditional Logit Analysis of Qualitative Choice Analysis, in P Zarembka (ed.): *Frontiers in Econometrics,* New York Academic Press, 105-42.

McClellan, M. (2011) Reforming Payments to Healthcare Providers: The Key to Slowing Healthcare Cost Growth while Improving Quality? *Journal of Economic Persp*ectives, Vol. 25, No. 2, 69-92.

Medical Professionalism Project (2002) Medical Professionalism in the New Millennium: A Physician Charter. Project of the ABIM Foundation, ACP-ASIM Foundation, and European Federation of Internal Medicine. *Annals of Internal Medicine*, Vol. 136, No. 3, 243-246.

Nilsen, S., Werner, E.L., Maeland, S., Eriksen, H.R., and Magnussen, L.H. (2011) Considerations Made by the General Practitioner When Dealing with Sick-Listing of Patients Suffering from Subjective and Composite Health Complains. *Scandinavian Journal of Primary Health Care,* Vol. 29, 7-12.

OECD (2010) Sickness, *Disability and Work: Breaking the Barriers. A Synthesis of Findings across OECD Countries*. OECD Publishing, Paris.

Pike, C. (2010) An Empirical Analysis of the Effects of GP Competition. MPRA (Munich Personal RePEc Archive) Paper No. 27613.

Santos, R. Gravelle, H. and Propper, C. (2017) Does Quality Affect Patients' Choice of Doctor? Evidence from England. *Economic Journal*, Vol. 127, No 600, 445–494.

Statistics Norway. (1999) Regionale inndelinger - En oversikt over standarder i norsk offisiell statistikk. Norges Offisielle Statistikk, Oslo-Kongsvinger.

Schwartz, L. (2002) Is there an Advocate in the House? The Role of Health Care Professionals in Patient Advocacy. *Journal of Medical Ethics,* Vol. 28, 37-40.

Wilkin, D. (1992) Patterns of referral: Explaining variation, in: M. Roland and A. Coulter, (eds.): *Hospital Referrals*, Oxford General Practice Series 22, Oxford University press.

Winde, L. D., Alexanderson, K., Carlsen, B., Kjeldgård, L., Wilteus, A. L., and Gjesdal, S. (2012) General Practitioners' Experiences with Sickness Certification: A Comparison of Survey Data from Sweden and Norway. *BMC Family Practice,* Vol. 13, No. 10

World Medical Association (2006) World Medical Association International Code of Medical Ethics.