MEMORANDUM

No 07/2009

The Gate is Open: Primary Care Physicians as Social Security Gatekeepers
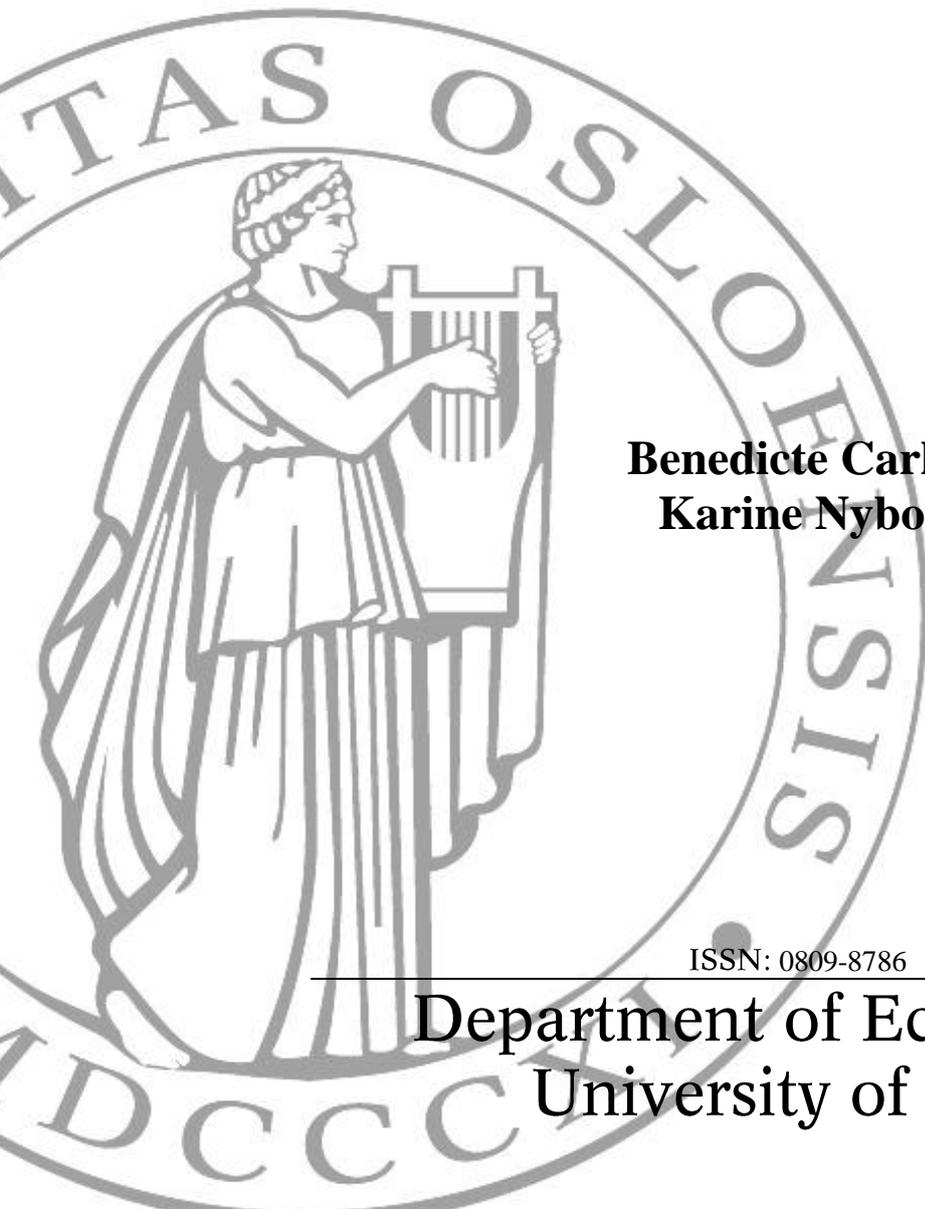
Benedicte Carlsen
Karine Nyborg

Department of Economics
University of Oslo

Last 10 Memoranda

Previous issues of the memo-series are available in a PDF® format at:
http://www.oekonomi.uio.no/memo/index.html

The Gate is Open: Primary Care Physicians as Social Security Gatekeepers

Benedicte Carlsen[1] and Karine Nyborg[23]

Abstract

Primary care physicians have two roles: the healer and the gatekeeper. We show that, due to information asymmetries, they cannot be expected to fulfill the latter role. Better gatekeepers will be poorer healers; hence all patients, both truly sick and shirkers, will strictly prefer physicians who give priority to healing. The choice between work and sick leave thus lies, essentially, with the patient. Interviews with Norwegian primary care physicians confirm this: Our interviewees report that shorter sick leaves are granted at request, while longer sick leaves are normally granted if the patient still prefer so after discussions with the physician.

Introduction

Most primary care physicians are supposed to fulfill two quite distinct roles; the healer and the gatekeeper. In welfare states with generous health-related social security benefits, like paid sick leave, the gatekeeping function seems particularly important.

Below, we will argue that unless primary care physicians have unreasonably strong preferences for being good gatekeepers, they cannot be expected to function as gatekeepers at all. The reason is that information asymmetries preclude physicians from distinguishing shirkers from certain groups of truly sick. As a consequence, there is an inherent conflict between the physician's two roles: being a better gatekeeper will always (expectedly) come at the cost of being a poorer healer. Moreover, if physicians compete for patients, this competition will amplify the conflict: all else given, a physician who gives priority to gatekeeping over healing will be preferred by *no* patient; hence such physicians will typically, in addition to lower incomes, get the satisfaction of being a good healer for only a small number of patients.

While the conflict between the healer and gatekeeper roles is widely acknowledged in the medical profession, the general consensus seems to be that when conflicts of interest arise, a physician's primary responsibility lies with the patient. For example, in the World Medical Association's *Code of*

Medical Ethics, physicians pledge that "the health of my patient will be my first consideration" (WMA 2006). Similarly, the *Charter of Medical Professionalism* (Medical Professionalism Project, 2002), states, as its first fundamental principle, the *primacy of patient welfare* – a dedication to serving the interest of the patient: "Market forces, social pressures, and administrative exigencies must not compromise this principle" (op.cit, p. 244). While these are normative rather than descriptive statements, several some studies find that physicians experience the dual role as a difficult balancing act. In his seminal work, Eisenberg (1986, p. 79) states: "Physicians seem to be most comfortable when they can consider only the welfare of their patients and themselves as the principal issues in medical care decisions [...] They are generally unwilling to consider the effect of their individual decisions on the overall cost of medical care or the consumption of limited resources (be they beds, scarce drugs, or dollars). In fact, many doctors believe that these broader considerations have no place in an ethical physician's decision making." More recent studies confirm that physicians find the gatekeeer role difficult (Carlsen & Norheim 2005; Pearson 2000; Arnesen & Fredriksen 1995; Ayres 1996; Angell 1993).

Information asymmetry in the patient – physician relationship has been studied extensively in the health economics literature. For example, researchers have explored the issue of demand inducement, i.e. the idea that physicians, for economic reasons, may recommend treatment which a perfectly informed patient would have viewed as excessive (e.g. Dranove 1988; Rochaix 1989; de Jaegher and Jegers 2000); optimal provider-physician contracts (e.g. Chalkley and Malcolmson 1998); and incentives for patient – physician collusion (Alger and Ma 2003). Our theoretical contribution lies in focusing specifically on the physician's dual roles as healer and gatekeeper, pointing out the inherent conflict between the two. For this purpose, we establish a model in which concepts such as health, symptoms, and their observability are spelled out in more detail than what is usually done in the health economics literature.

We show that under plausible conditions, physicians can only keep shirkers away from unjustified social security benefits at the cost of becoming poorer healers and, in addition, facing a substantial reduction in patient demand. Unless physicians have very strong intrinsic preferences for being good gatekeepers (to an extent appearing rather unreasonable given the professional attitudes quoted above), they will not function as gatekeepers at all: The gate is open. This holds even under the assumption that when physicians are confident about the true diagnosis, they sicklist only if a sick leave is required as part of the medically recommended treatment.

Such a result, however, leaves open the issue of why some workers still work. In Norway, sick leave for up to one year is fully compensated, and it is illegal to fire employees because of health problems. In addition, most of the costs of seeing a doctor is covered by the public sector; patients pay only a modest fee. Under these conditions, and with open gates, an economist would expect shirking to be extremely widespread. Norwegian policy makers do worry about relatively high sick leave rates and the associated costs; however, sick leaves are still substantially lower than one might expect under the prevailing circumstances: On a typical working day, about 6 - 7 percent of Norwegian workers are absent due to sickness (Markussen et al., 2008). This can have at least two potential explanations: First, there may be private benefits of being present at work (or not being sicklisted), over and above the contracted wage, that make workers prefer not to be sicklisted. For example, working may involve benefits in terms of social interaction, it may improve one's self-image as important to

others, or be intrinsically interesting; in addition, even with full wage compensation there may be pecuniary incentives, for example because employers may interpret absence as a signal of either poor health or poor motivation, reducing chances of promotion and increased future income.

The second possible explanation is that our theoretical analysis simply is wrong, and that physicians do function as gatekeepers after all. To explore this possibility one step further, we present data from focus group interviews with Norwegian primary care physicians. Our informants confirm that they hardly ever decline patients' requests for short-term sickness certificates. In the case of long-term sick leave, they claim to be slightly more active, using various strategies to persuade patients that it would be better for them to go back to work. Even in these cases, however, physicians 'negotiate' with patients (a term commonly used by our informants) rather than directly rejecting appeals for sick leave; and if they do not succeed in convincing the patient that returning to work in his or her own interest, a sickness certificate will typically be granted or prolonged. The findings in this part of our study are in line with several other Scandinavian surveys, indicating that primary care physicians rarely deny sickness certificates (Wahlström and Alexanderson 2004; Carlsen and Norheim 2003; Gulbrandsen, Førde and Aasland 2002; Englund, Tibblin and Svardsudd 2000; Larsen, Førde and Tellnes 1994).

Our informants claim that most sick leave certificates are clearly indicated on medical grounds and thus not questionable. Moreover, even among the questionable cases, physicians in our study did not believe that much cheating on patients' part was taking place. Obviously, if a patient succeeds in cheating her physician, the physician will not be aware of this, so one should perhaps not attach too much weight to these statements. Nevertheless, given the very modest gatekeeping activities reported by our informants, it is somewhat surprising that they do not suspect, to a larger extent, that patients come to see them with the aim of exploiting the system.

In our view, these results strengthen the case for our first proposed explanation: Most workers seem, for reasons left open in the present paper, to actually prefer work to paid sick leave.

Healing vs. gatekeeping: a formal model

Health

For simplicity, assume that each worker lives for two periods. At the beginning of period 1, a vector of k>1 possible symptoms, $S_i = \{\sigma_i^1,..., \sigma_i^k\}$, is drawn by nature for each worker *i*. Let $\sigma_i^j \in \{0,1\}$, where 1 denotes that the symptom is present, while 0 means that the symptom is absent. Symptoms $\sigma_i^1,..., \sigma_i^g$ (where 0<g<k) are observable only by a physician (e.g. blood test results, conclusions from ultrasound images); we will call these *verifiable* symptoms[4]. Symptoms $\sigma_i^{g+1},..., \sigma_i^k$, on the other hand, are observable only by the patient herself (e.g. pain, nausea, dizziness, blurred vision, mental disturbances). We will call these *subjective* symptoms.[5]

---

[4] Strictly speaking, using the term "symptom" about σ_i^1 might be appropriate only when $\sigma_i^1=1$, since otherwise there is no symptom; here, we will sometimes also use the term "symptom" for a variable such as σ_i^1 independently of its value.
[5] For simplicity, we disregard symptoms observable by both parties (e.g. a skin rash), or by no-one. Further, we assume that there is a large enough number of possible, and sufficiently frequent, symptom combinations that we can neglect the possibility that a patient correctly guesses her own vector of verifiable symptoms before seeing a doctor.

Denote by h_i^t patient i's subjectively experienced general health status at time t. h_i^t as such is only observable by the patient herself. Let $h_i^t \in \{0,1\}$, where 0 means that the patient feels sick, while 1 means that she feels well .

Some combinations of symptoms will correspond to a good subjective health status ($h_i^t =1$), while other combinations of symptoms will correspond to sickness ($h_i^t =0$). Assume that for every set of symptom variables $S_i^t = \{\sigma_i^1,..., \sigma_i^k\}$, there exists at most one correct diagnosis $D_i = D(S_i)$ (where "healthy" is one possibility).

Arriving at a diagnosis will generally require information both from the physician and the patient. However, in some cases the diagnosis may depend on only one party's observations. We will call a diagnosis $D(S_i)$ *subjective* if all indicating symptoms are observable by the patient only. Depression, severe back pain or headaches, or chronic fatigue syndrome may possibly be examples of such diagnoses.[6] A diagnosis is said to be *verifiable* if information observable only by the physician is required to establish the diagnosis. The diagnosis is *strictly verifiable* if the physician can arrive at the diagnosis based on her own observations only (the diagnosis is unaffected by subjective symptoms). A diagnosis is *weakly verifiable* if observations from both parties are required to arrive at the diagnosis.

At the beginning of period 1 (but after health status for that period is drawn), the patient may decide to see a physician. The patient's cost of doing so is disregarded for simplicity, and can be considered covered by the government. Adding a fixed cost of seeing the doctor would not change the results below substantially, except that healthy workers would require a somewhat stronger utility gain from sick leave before preferring to see a doctor. If the patient is indifferent, we will assume that she does not see a physician.

For every diagnosis we assume that there is one and only one associated correct recommended treatment $T(D_i)$. One element of this treatment may be a maximum recommended working time $L(D_i)$. There may also be a minimum recommended working time; for simplicity, we disregard this here.

Assume that every individual has the same contracted working hours τ. If $L(D_i) < \tau$, the patient is entitled to a sickness certificate for period t. If a patient is sicklisted in period t, she is sicklisted for the entire period. There will generally also be other elements of the recommended treatment, which we may think of as medication (or radiation treatment, physiotherapy, etc.), which may require a prescription, referral, or just information from the physician. Only physicians have full knowledge of the steps from symptoms to diagnosis and from diagnosis to recommended treatment.[7]

In the second period, person i's health status h_i^2 is assumed to be identical to h_i^1, with the following important exception: If $h_i^1=0$ and person i receives adequate treatment $T(D_i)$ in period 1, there is a probability $0<q<1$ that the person will be perfectly healed ($h_i^2=1$). In other words, if the worker is sick

---

[6] A diagnosis is subjective if, for a given set of subjective symptoms variable values $\{\underline{\sigma}_i^{g+1},..., \underline{\sigma}_i^k\}$, the diagnosis $D(S_i)= D(\{\sigma_i^1,..., \sigma_i^g, \underline{\sigma}_i^{g+1},..., \underline{\sigma}_i^k\})$ is unaffected by the values taken by the verifiable symptom variables.

[7] We assume that the number of possible treatments, as well as the frequency with which each is required, are such that we can disregard the possibility of correct treatments being prescribed by chance.

in the first period, the only way to get well is to get adequate treatment, but even then there is a probability that the treatment will not be effective. If one is well to begin with, one stays well.

Let us focus on what happens in the first period. If a person receives treatment $T(D_i)$, and if, moreover, $L(D_i) < \tau$, this is taken to mean that person i is sicklisted and also gets the correct medication. We do not model medication explicitly since it has no substantial impact on the analysis; its role here is simply to ensure that the patient cannot treat himself without seeing a doctor simply by staying home from work: the physician's expert knowledge is needed.

Expected health consequently develops over time as follows:

(1) $E(h_i^2)$ $= 1$ if $h_i^1 = 1$

$= q$ if $h_i^1 = 0$ and $T_i = T(D_i)$

$= 0$ if $h_i^1 = 0$ and $T_i \neq T(D_i)$.

Work and workers

Let us now turn to the incentives of the agents: workers and physicians.

Consider, first, workers. (If she decides to see a physician, a worker will also be considered a patient, whether she is truly sick or not.) Assume that workers maximize expected utility, where utility is given by

(2) $U = U_i^1 + U_i^2$

where

(3) $U_i^t = (\alpha c_i^t + \beta l_i^t + \gamma_i L_i^t) h_i^t$

Here, $\alpha > 0$, $\beta > 0$, while $\gamma_i \geq 0$. c_i^t is i's consumption in period t, l_i^t is her leisure time in period t, while L_i^t is the time i is present at work in period t. Linear separability in consumption, leisure and working time is assumed for simplicity; for the same reason, we disregard discounting. Note that with this specification of utility, nothing is fun when you are sick; thus health is always the worker's first priority.

If $\gamma_i > 0$, the worker gets a private benefit from being present at work (whenever her health is good). This may reflect the pleasure from social interaction in the workplace, self-image benefits from doing a good job/being important to others (Brekke and Nyborg 2008), the satisfaction of performing intrinsically interesting tasks, or the pleasure of developing one's skills. It could also reflect that the probability of promotion (renegotiation one's labor contract in period 2 and succeeding in getting a higher wage) may be higher if the worker is present at work in period 1 (Markussen, 2009). To keep the model as simple as possible, however, we will assume below that wages are exogenously given, implying that any private benefits of being present at work will be treated as if they were non-pecuniary; this restriction is not essential, however. The distribution of individual work motivation γ_i is assumed to be independent of the distribution of symptoms S_i.

Labor contracts are entered in the beginning of period 1, before S_i is drawn. With the assumptions of the present paper, all workers will prefer to have labor contracts, hence we will simply assume that all workers have a labor contract from the beginning of period 1.

A labor contract is a binding contract between employer and employee, specifying that the employee is to provide τ working hours per period in return for an exogenously fixed wage for each period, w_i^t. The employer is entitled to cancel the contract after period 1 only if the employee has provided less than τ working hours in period 1 without presenting a valid sickness certificate.[8]

Any patient receiving a sickness certificate is entitled to full wage compensation; for the present analysis, it will not matter whether sickness benefits are paid by the employer or the government.

The worker's time constraint in period t is

(4) $1 = l_i^t + L_i^t$

where total time is normalized to 1. Furthermore, the relationship between working time L_i^t and sicklisted time s_i^t is given by

(5) $L_i^t = \tau - s_i^t$

Disregarding credit markets, the individual's budget constraints in period 1 and 2, respectively, provided that she keeps her labor contract, are

(6) $w_i^1 = c_i^1 + C_i$

(7) $w_i^2 = c_i^2$

where C_i is medical treatment cost, if any, paid by i in period 1.

We will focus on the following three choices on the patient's part: First, whether or not to see a physician; second, which physician to see; and third, whether or not to represent subjective symptoms truthfully to the physician.

First, note that a sick worker ($h_i^1 = 0$) enjoys nothing in period 1; her first priority is thus to heal and become well in period 2. Her only chance of doing so is to get the correct treatment, which means staying away from work if a sick leave is recommended for her diagnosis, as well as taking the required medication. Since only physicians know the relationships between symptoms and recommended treatments, and patients cannot guess their own diagnosis, a sick worker will choose to see a physician, irrespective of her wages w_i^t and her work motivation γ_i.[9] If the patient has reason to believe that her probability of getting the correct treatment differs between doctors, she will seek those doctors for whom this probability is highest; that is, the best healers. Moreover, to maximize

---

[8] It is natural to assume that the employee is free to quit before period 2 if she prefers to do so; however, since we will not explicitly analyze contract renegotiation, this will not be of importance.
[9] Note that if sick leave was the only component of medical treatments, a sick worker could treat herself simply by staying home from work in period 1 (breaking her employment contract if necessary). She would then need the physician only for economic purposes, that is, for getting paid while absent and not having to quit her job. The presence of diagnosis-specific medical treatment in the model ensures that the physician's expertise is needed (in a non-trivial way) for healing.

the probability of getting the correct treatment, she will present her subjective symptoms accurately to the doctor. Unless the recommended treatment cost exceeds her income w_i^1, which we will disregard in the following, she will undergo the treatment recommended by the physician.

For a healthy worker ($h_i^1=1$), things are different. If a healthy worker is sicklisted, there is no health gain; there is, however, a benefit in terms of increased leisure. In addition, there may be a cost in terms of lost private (non-wage) benefits from being present at work. In the absence of a health gain, sicklisting in period 1 has no effect on utility in period 2. Hence, for healthy workers, utility in each period can be considered separately. We assume that a healthy worker never pays treatment costs (if the physician has prescribed medication, the patient can simply abstain from taking it).

A healthy worker will prefer to be sicklisted in period 1 if and only if the benefits of extra leisure exceed the costs of not being present at work. Let $\Delta U_i^1(h_i^1=1)$ denote the utility gain from being sicklisted in period 1 for a healthy worker (as compared to working τ hours). Using that $l_i^1 = 1 - \tau + s_i^t$ this can be written

(8) $\Delta U_i^1(h_i^1=1) = \tau(\beta - \gamma_i)$

Hence, the gain of being sicklisted is strictly positive if and only if

(9) $\gamma_i < \beta$

that is, if the individual's marginal benefit of working is lower than her marginal value of leisure.

A shirker reporting symptoms corresponding to a verifiable diagnosis would require sheer luck to get a sickness certificate, since the cannot foresee the physicians' observations. Below we assume that the probability of this can be neglected. Assume, however, that *each worker knows the symptoms of at least one subjective diagnosis requiring sick leave*. If a shirker reports to have these subjective symptoms, no information whatsoever is available for the physician to distinguish the shirker from truly sick patients with the very same symptoms. Provided that there is a strictly positive probability that physicians will grant sickness certificates for subjective diagnoses, shirkers will thus falsely report to have symptoms of a subjective diagnosis.

Consequently, we can divide workers into three groups, with associated optimal strategies and numbers:[10]

  A. *Sick workers:* Workers who are sick in period 1 ($h_i^1=0$) will always see a physician; they will seek those physicians who are thought to be the best healers, and they will never misrepresent their subjective symptoms. This holds regardless of their work motivation.

  B. *Motivated, healthy workers:* Healthy workers ($h_i^1=1$) with high work motivation ($\gamma_i > \beta$) will prefer not to be sicklisted. These workers will never see a doctor.

---

[10] A fourth category, which is mentioned by physicians in our empirical study, is the sick worker who insists on working, refusing a sick leave (called *presenteeism* in the literature, see Aronsson and Gustafsson (2005). This category is excluded in the theoretical model because our assumptions ensure that workers always give priority to health concerns over other issues.

C. *Shirkers:* Healthy workers (h_i^1=1) with low work motivation ($\gamma_i < \beta$) will prefer to be sicklisted. If there is a positive probability that the doctor will grant a sick leave, these workers will see a doctor; they will seek those physicians who are thought to be the poorest gatekeepers; and, provided that there is a positive probability that the physician will grant sick leave for subjective diagnoses, they will falsely report to have symptoms corresponding to a subjective diagnosis requiring sick leave.

Let N be the total worker population, σ the share of the population who are truly sick, and γ the share of workers with $\gamma_i > \beta$. Then, using the terminology from A – C above, $N\sigma$ workers will (expectedly) be sick, $N\gamma(1-\sigma)$ will be motivated and healthy, and $N(1-\sigma)(1-\gamma)$ will be shirkers.

Physicians

The physician's job is to establish a diagnosis, and on the basis of this diagnosis recommend treatment, including sick leave. Our focus here will be on the decision to grant or not grant a sickness certificate. For expositional convenience, assume that all subjective diagnoses require a sick leave. Assume also that the patient is always asked to report her subjective symptoms to the physician *before* the physician reveals the verifiable symptoms to the patient. Let us first discuss the problems faced by the physician in arriving at a diagnosis and thus the correct recommended treatment; then, we will turn to the physician's preferences, incentives, and optimal strategies.

If a patient displays the symptoms of a strictly verifiable diagnosis, the physician's task is straightforward. She can then simply make her own observations, establish the true diagnosis $D(S_i)$ based on this, and grant a sick leave if and only if $L(D(S_i)) < \tau$. If a patient displays the verifiable symptoms of a weakly verifiable diagnosis, the physician needs information from the patient to arrive at the correct diagnosis. In this case, however, there is little reason to suspect the patient to be misrepresenting her symptoms; for a shirker it is a better strategy, as explained above, to report the subjective symptoms of a subjective diagnosis she is familiar with (e.g. depression, back pain).

The physician can thus reasonably conclude that there will not be shirkers among patients who present symptoms of verifiable diagnoses. The choice of whether or not to sicklist can, in these cases, be based exclusively on medical criteria, that is, whether or not $L(D_i) < \tau$. In these cases there is no inherent conflict between the gatekeeper and healer roles.[11]

It is not reasonable to expect all patients with subjective diagnoses to be honest. However, among these patients, there is no available information that can help sorting honest from dishonest patients. The decision of whether or not to grant a sick leave then amounts to deciding whether or not to accept patients' information at face value.

The physician basically has three possible strategies in the case of subjective diagnoses. First, she can act as if such patients are always telling the truth, base her diagnosis on this, and grant a sick leave. Let us call this strategy *Lax*. Second, she can act as if such patients are always lying, that is, never grant a sick leave in the case of subjective diagnoses. We will call this strategy *Strict*. Third, she can use any combination of these two strategies, by granting sick leave for only some of the patients

---

[11] The physician might, of course, deviate from the medically correct sicklisting practice even in these cases if there was an economic incentive to do so. We will get back to physicians' incentives below.

reporting symptoms of subjective diagnoses. Note, however, that since no information is available on which to base such differentiation, the third type of strategy, which we may call Mixed, boils down to pure randomization.[12]

Let us now turn to physicians' incentives and optimal strategies. Assume that each physician's labor supply is exogenously given. Physicians care about their monetary income; in addition, we will allow that a physician cares about doing a good job, which has two aspects: success as a healer, and success as a gatekeeper. Thus, let physician p's utility u_p in any period (we omit superscripts for periods, since they will not matter) be

(10) $u_p = Y_p + b_p H_p + g_p G_p$

where Y^p is physician p's income, H_p is the physician's belief about her success as a healer, G^p is the physician's belief about her success as a gatekeeper, $b_p \geq 0$ and $g_p \geq 0$. Linear separability is assumed for the sake of simplification. We assume, moreover, that there exists at least one physician for whom $g_p = 0$.

Let physician p's income be proportional, by a factor $a > 0$, to her number of patients, N_p:

(11) $Y_p = aN_p$

The physician's beliefs about her success as a healer, H_p, depends on her beliefs about the health effects of her own work. For every patient getting a correct treatment, the physician knows that the healing probability is q; hence her belief about her success as a healer must somehow depend on the number of patients the physician believes she has treated correctly.

To avoid complicating things unnecessarily, let us assume that when there is no medical doubt, i.e. when the physician is confident about the true diagnosis, she sicklists if and only if a sick leave is a part of the recommended treatment. This means that the physician will always treat patients with verifiable diagnoses, and expect this to yield a health gain of q per patient. For patients with subjective diagnoses, the expected health gain will depend on the physician's belief about the share of truth-tellers among these patients, and on whether she follows a Lax or a Strict strategy.

Let L_p denote the laxness of physician p's strategy, where $L_p = 1$ means Lax and $L_p = 0$ means Strict.[13] Further, let N_{pV} be the number of p's patients who have verifiable diagnoses, while N_{pS} is the number of p's patients who report symptoms of subjective diagnosis (such that $N_p = N_{pV} + N_{pS}$). We will then specify p's belief about her success as a healer as

(12) $H_p = q(N_{pV} + L_p \theta_p N_{pS})$

where θ_p is p's belief about the share of truth-tellers among her patients claiming subjective diagnoses.

---

[12] That is, any physician has two possible pure strategies, Lax and Strict; any combination of these two will constitute a mixed strategy. Below, we will largely disregard the mixed strategies, since they do not play any crucial role in our main results.

[13] Mixed can be indicated by letting L_p take a number between 0 and 1, depending on the probability of sicklisting in p's strategy.

Let us further define being a good gatekeeper as *not granting unjustified sickness certificates*, and, correspondingly, specify G_p as decreasing proportionally with the physician's belief about the number of unjustified sickness certificates she grants.[14] Thus,

(13) $G_p = - L_p(1-\theta_p)N_{pS}$

For patients with subjective diagnoses, there is an inherent conflict between the physician's healer and gatekeeping roles: A Lax physician will be a strictly better healer than a Strict physician, since the Strict does not even attempt to treat these patients. The Lax will, however, also be poorer gatekeepers, since they do not even try to stop shirkers. Note also that this dilemma cannot be solved, or even mitigated, by choosing Mixed; like Strict, a Mixed strategy makes the physician a strictly poorer healer, but also a strictly better gatekeeper, than if Lax. Since there is no way to distinguish truth-tellers from liars in the case of subjective diagnoses, stricter gatekeeping will always produce strictly lower expected health benefits. It is simply not possible to improve one's gatekeeping success without, at the same time, reducing one's expected success as a healer.

Let us now consider what determines the number and types of patients coming to see a given physician. Assume that practice styles Lax and Strict are observable to patients. This seems reasonable; for example, previous patients may have conveyed information about the physicians' sickness certificate decisions in their cases.[15]

Above, we established that *motivated, healthy workers* will never see a doctor; *sick workers* will see a physician, seeking the best healers; while *shirkers* will see a doctor, seeking the poorest gatekeepers. Having established that the Lax doctors are both the best healers and the poorest gatekeepers, we can thus conclude that *every patient* (that is, every worker who does see a doctor) *strictly prefers a Lax physician*.

Let us explain this in some more detail. For anyone with a verifiable diagnosis, any doctor will, in fact, be equally good; all doctors will treat these patients correctly. However, recall that the public does not generally have medical knowledge. For this reason, the truly sick patient will typically not know, before seeing a doctor, whether her diagnosis is, in fact, subjective or verifiable; she knows only that she does not feel well ($h_i^1=0$), but does not know which symptoms are relevant to establish a diagnosis in her particular case. Since there is typically a positive probability, from the patient's point of view, that her diagnosis is of the subjective kind, she will strictly prefer a Lax physician. The only type of patient who knows, in advance, which type of diagnosis she has, is the patient who happens to have (one of) the subjective diagnoses that she is actually familiar with. These patients, however, will obviously prefer Lax physicians. Hence, *all* patients will strictly prefer Lax to Strict physicians. Moreover, since the Mixed strategy is just a combination between Lax and Strict, all patients will also strictly prefer Lax physicians to Mixed.

---

[14] Hence one may maximize gatekeeping success simply by granting a sick leave to nobody, regardless of health status. An alternative measure of gatekeeping success would be to minimize the number of errors, where a granted, but unjustified, sick leave would produce the same disutility as a denied, but justified, sick leave.

[15] If the worker knows one patient who has been denied sick leave for a subjective diagnosis, she will know that the physician is either Mixed or Strict. A possible disability to distinguish Mixed from Strict will not be essential here, though, since patients will always strictly prefer Lax physicians to both Strict and Mixed.

The above conclusion will obviously tend to secure a higher demand, and thus a higher income, for Lax physicians than for Strict or Mixed. In addition, however, a physician's number of patients will also affect her success as a gatekeeper and her success as a healer. Let us now turn to discussing the market equilibrium.

Market equilibrium

Let us now consider possible Nash equilibria in this market: given everybody else's strategies, each patient's and physician's strategy is individually optimal.

Note first that due to the assumption that there exists at least one doctor with $g_p = 0$, there will always be at least one Lax physician. Further, note that workers' choice of whether or not to see a physician, as well as their preference for Lax over Strict physicians, are then independent of others' choices: Workers who are sick or who prefer a sick leave will want to see a physician, and will prefer one who is Lax, regardless of other workers' or physicians' strategies.

The total number of patients in the economy equals $(1-\gamma+\sigma\gamma)N$. If a patient is indifferent between two physicians, she is assumed to choose randomly between them. Disregarding the Mixed strategies, which do not add anything substantial, there are three possible situations concerning the number of patients faced by physicians of different strategies:

1) No capacity constraints are binding; Lax physicians can handle all demand. Lax physicians will then share the total number of patients equally, while Strict physicians get no patients at all.

2) Physicians have a capacity constraint, which is binding for the Lax physicians, but not for the Strict.[16] Strict physicians will then get a positive number of patients, although strictly fewer than the Lax.

3) All physicians, Lax and Strict, fill their entire capacity. In this case all physicians get the same number of patients.

Case 3 implies that physicians' services are rationed to such an extent that patients' physician preferences play no role at all; thus, this case is hardly the most interesting. Concerning Case 2, note that – unless the situation is very close to Case 3 in the sense that Strict physicians are close to their capacity constraint – Strict physicians will have a substantially lower number of patients than Lax. There are two reasons for this. First, no Strict physician gets any patient at all until all Lax physicians have filled their capacity. Second, the number of patients shared between the Strict when Lax have filled their capacity will be strictly lower than the number of patients declined by Lax physicians, because shirkers, in addition to patients who know their own subjective diagnosis, will not bother to see a Strict doctor at all.[17]

Case 1 is the simplest to analyze. In Case 2, results are modified compared to Case 1; still, the mechanisms at work will be relatively similar. Let us thus focus on the simplest case where physicians do not face capacity constraints.

---

[16] We assume that this constraint is the same for all physicians, independently of laxness strategy.
[17] Note, though: they *would* bother to see a Mixed doctor.

Let $N_p(L_p)$ denote the number of patients faced by physician p given that her strategy is L_p (recall that $L_p=1$ means Lax and $L_p=0$ means Strict), and use a similar notation for patients with verifiable or subjective diagnoses, respectively (so that $N_{pV}(L_p)$ is the number of patients with verifiable diagnoses faced by *p*). We then have that

(14) $N_p(L_p) = ((1-\gamma+\sigma\gamma)N\, L_p\,)/(\sum_p L_p)$

where $\sum_p L^p$ is the total number of Lax physicians.

The utility of a physician will be a function of her practice strategy. Inserting from above, we get

(15) $u_p(L_p) = aN_p(L_p) + b_p q(N_{pV}(L_p) + L_p\theta_p N_{ps}(L_p)) - g_p L_p(1-\theta_p)N_{ps}(L_p)$

Physicians have no way to distinguish any given shirker from any given truly sick patient with a purely subjective diagnosis. Nevertheless, assume that generally available information such as rumours, general talk, anonymous survey statistics and so forth implies that the *aggregate* shares of motivated workers, γ, and sick workers, σ, are publicly known with reasonable accuracy. Since Lax doctors now share all demand between them, there is no reason to expect the composition of p's patients to differ from the aggregate composition of the patient population. Let us denote by π the aggregate share of shirkers in the patient population (those who actually see a physician). π will then be given as

(16) $\pi = [(1-\gamma)(1-\sigma)] /[\sigma+(1-\gamma)(1-\sigma)]$

For the Lax doctors, this implies that eq. (15) can be written (where $N_p(1)$ is given by eq. (14)):

(17) $u_p(1) = N_p(1)[a + b_p q\, (1-\pi) - g_p\pi]$

using that the expected share of shirkers among p's patients equals π. For the strict doctors, we have that $N_p(L_p) = N_{pV}(L_p) = N_{ps}(L_p) = 0$, implying that $u_p(0) = 0$. Thus, in the case with no binding capacity constraints, a physician *p* will prefer to be Lax whenever $u_p(1) > 0$.

$u_p(1)$ does depend on other physicians' strategies, since $N_p(1)$, the number of patients faced by *p*, is higher the lower the total number of Lax physicians. Nevertheless, since $N_p(1) > 0$, the size of this number will only scale the entire utility level of a Lax physician *p* upwards (if $u_p(1) > 0$) or downwards (if $u_p(1) < 0$); it will not affect the decision of being Lax or Strict. The latter decision is determined by the sign of the term in square brackets in eq. (17), which determines the sign of $u_p(1)$. Hence, *p* will prefer to be Strict if and only if

(18) $a + b_p q\, (1-\pi) - g_p\pi < 0.$

There are several things to be noted here. First, if $g_p = 0$, there is no incentive at all to be Strict. If $g_p > 0$, there is such an incentive, and its size depends on the aggregate share of shirkers among patients. Second, the desire to be a good healer provides an incentive *not* to be a good gatekeeper, since gatekeepers are poorer healers. Note that the only patients who would, in fact, receive poorer treatment from a Strict physician are those who are truly sick with subjective diagnoses; but in (18), the healing benefit $b_p q$ is multiplied by $(1-\pi)$, which includes all non-shirking patients, even those with verifiable diagnoses. The reason is the competition for patients in the market for physician

12

services: In addition to making one a poorer healer for patients with subjective diagnoses, being Strict also scares away all one's other patients, depriving the physician of the opportunity to heal those as well.[18]

Third, and again due to market forces, the income per patient a is multiplied by 1 in eq. (18): By being Strict, the physician will lose all her patients, not only that share of them who claim subjective diagnoses and would thus have been treated differently.[19]

Finally, note that since the first term in (18) is strictly positive, then even if all patients were known to be shirkers (π=1), a positive gatekeeping preference would not be sufficient to secure that the physician is Strict: In fact, even in this extreme case, physician p will only prefer to be Strict if $a < g_p$ (i.e., the income from one extra patient counts less in p's utility than the satisfaction of avoiding one false sickness certificate).

Eq. (18) implies that for every physician p, there will be a threshold π_p* such that if the aggregate share of patients who are shirkers exceed this threshold, p prefers to be Strict:

(19) π_p* = [a + b$_p$q]/ [g$_p$ + b$_p$q]

On the right hand side of eq. (19), the numerator consists of the benefit to physician p of treating a truly sick patient (consisting of income plus the satisfaction of being a good healer). The denominator is the *utility loss* compared to this if the treated patient is, instead, a shirker: The physician gets the per patient income anyway, but does not get the satisfaction of healing the patient, and will, in addition, experience the dissatisfaction of being a poor gatekeeper. From (19) it can easily be seen that if a > g$_p$, we would get π* ≥1, which means that (18) can never hold; hence if a > g$_p$, p will always be Lax, regardless of other physicians' strategies and of the aggregate share of shirkers.

If a < g$_p$, is it possible for π* to take values between 0 and 1, so in this case the physician will be Strict if the aggregate share of shirkers is sufficiently low. Note, however, that a < g$_p$ is a necessary but not sufficient condition for the equilibrium existence of Strict doctors. The benefits of being a good healer, measured by b$_p$q, appears both in the numerator and denominator of (19); hence, if p primarily cares about being a good healer, π_p* will be close to 1 regardless of the relative size of *a* and g$_p$, and p will only be Strict if the share of shirkers is very high.[20] More generally, even if we disregard the economic incentive, a physician who does care about gatekeeping, but who cares considerably more about healing, will prefer to be Lax – disregarding her gatekeeping role altogether – unless she has reason to believe that the patient population for the most part consists of liars. The economic incentive will, of course, reinforce this conclusion.

---

[18] Let σ^k be the share of workers sick with diagnosis type k, such that $\sigma^s + \sigma^v = \sigma$. Further, let s = σ^s/[σ+(1-γ)(1-σ)] be the share of patients truly sick with subjective diagnoses, and v = σ^v/[σ+(1-γ)(1-σ)] be the share of patients sick with verifiable diagnoses, so that 1 = π + s + v. Then, a share s < (1- π) of patients will have lower expected health gains if treated by a Strict physician. The physician, however, will lose all his patients, including those who would not be affected (since patients themselves do not who these are).

[19] Using the same notation as in the previous footnote: If one expected that only shirkers and patients sick with subjective diagnoses would flee Strict doctors, the income incentive would amount to a(π + s), which is strictly less than a.

[20] For example, assume that a=0 so that economic incentives are absent altogether, that q= 0.6, and b$_p$ = 5 and g$_p$ = 1. This yields π* = 0.75, implying that p will be Lax unless she has reason to believe that less than 25 percent of all patients are truly sick.

Thus, we can conclude that if $a > g_p$ for all p, the only Nash equilibrium is the case where all physicians are Lax, while all sick workers and shirkers see a Lax physician.

The situation where all physicians are Strict can never be a Nash equilibrium, since there will, by assumption, always be at least one physician with $g_p = 0$ who will thus prefer to be Lax. If $a > g_p$ for some p, however, it is possible that there exist Nash equilibria with a strictly positive number of Strict physicians; this will depend on the share of shirkers in the population. These Strict physicians would, however, get no patients. Given the attitudes in the medical profession, we do not find this kind of Nash equilibrium very plausible.

The above analysis assumed that there were no binding capacity constraints. If that is not the case, the results will obviously be modified. In particular, the Strict would face a strictly positive number of patients, but strictly fewer, and possibly substantially so, than the Lax doctors. The Strict will still be perfect gatekeepers; in addition, they will also get some income and healing benefits. Nevertheless, both their income and healing benefits will typically be considerably lower than those of the Lax doctors, simply because, unless they too are close to their capacity constraints, they have considerably fewer patients. Consequently, even if there is some excess demand left for Strict physicians to handle, the main conclusion will still hold: a physician needs to have a very strong preference for gatekeeping to choose being Strict.[21]

With plausible assumptions, thus, all physicians will be Lax in the market equilibrium; moreover, if there are Strict physicians, they will be preferred by no patients. The key to understand these dramatic results is that being a good gatekeeper – being Strict – does not only have costs in terms of the lost healing probability for one's patients with subjective diagnoses. It also leads to a substantial loss of other patients – both shirkers and patients with verifiable diagnoses. Consequently, the costs of being Strict are considerable; in addition to the income loss, one loses the opportunity to be a good healer for all these other patients as well.

One might modify the above model by assuming that there exists some imperfect signal indicating that the patient is lying: she may, for example, be blushing. This would open up the possibility of strategies of the following type: Of the patients with subjective symptoms, treat only those who do not blush. As long as the signal is imperfect, however, using such a strategy will make the physician a strictly poorer healer (expectedly) than if Lax; and since patients will know this, substantial parts of the above analysis will follow through as before.

The conclusions thus far may seem to leave rather bleak prospects for a generous sick leave system: unless physicians have very strong preferences for gatekeeping, they cannot be expected to function as gatekeepers at all. This holds even if physicians never grant a sick leave when confident that the patient does not need one. Among the two non-pecuniary motives, healing and gatekeeping, physicians' professional training obviously supports the first one much more strongly than the second. Since the pecuniary motive also works against the gatekeeper role, it seems very unlikely that many physicians will prefer being Strict.

---

[21] Note that Strict doctors will only have truth-telling patients. Knowing this, one may think that they would change their strategy and become Lax; however, as soon as they do so, they will attract shirkers, which will change their patient population. For the choice of practice style, it is thus still the aggregate share of shirkers that is relevant.

Policy and welfare effects

A full welfare economic analysis is beyond the scope of the present paper, but some remarks are in order.

There are two main reasons why physicians are reluctant to be gatekeepers in the above analysis. The first originates from the inherent conflict between healing and gatekeeping in all cases of patients reporting symptoms of subjective diagnoses, which, due to limited information, makes better gatekeeping impossible except though becoming a poorer healer. The second problem is of a slightly different kind: since sick people typically do not know whether their diagnosis is a subjective one, and since shirkers will not bother to see a Strict physician anyway, being Strict will deprive the physician of all, or at least a large share, of her demand from patients.

While the second problem is related to the market structure and physicians' incentives, and might thus potentially be solved by policy intervention, the underlying basic problem is the first, i.e. the inherent conflict between the gatekeeper and healer roles. Since this conflict is caused by an information problem, it is not obvious how it can be amended by public policy measures. A regulator may be able to change physicians' preferred policy by introducing additional incentives, but can hardly resolve the inherent conflict between gatekeeping and healing.

More specifically, the regulator may for example introduce a fee for sicklisting, implying that the per patient income to physicians is lower for sicklisted than for not sicklisted patients. This would make it more profitable to be Strict. If all physicians, as a consequence of a such fee, changed their practice styles from Lax to Strict, while still issuing sick leaves according to medical criteria in cases of verifiable diagnoses, the number of unjustified sick leaves would be reduced from $N\pi$ to zero; aggregate health benefits would also be reduced, by $Nq\sigma^s$ (where σ^s is the share of the worker population suffering from subjective diagnoses). Note, however, that the lost health benefits affect only that share of the population who are truly sick with subjective diagnoses. Individual physicians, in their choice between Strict and Lax practice styles, face an even harsher conflict between their healer and gatekeeper roles: they must also take into account that if Strict, they will lose most of their patients. That is, in addition to the real loss of health benefits, there is a distributional effect: The benefits of being a good healer for those with verifiable diagnoses go to the Lax, not to the Strict, physicians.

Would ensuring that all physicians were Strict rather than Lax be a net welfare improvement? This would depend on the initial share of shirkers, and of truly sick with subjective diagnoses, as well as the social importance attached to health benefits and to reduced gatekeeping failure. While the lost health benefits obviously represent a real cost, the unjustified sick leave benefits paid to shirkers will partly be a pure transfer, not a social cost; to which extent will depend, if these benefits are covered by the public sector, on the social cost of funds.[22]

A fee on physicians' issuing of sickness certificates, however, would also introduce new incentives for physicians to deviate from the medically justified conclusions even in verifiable cases. With a sufficiently high fee, some physicians would be tempted to deny sickness certificates even in cases

---

[22] A full welfare analysis would be slightly more complex; it would, among other things, have to include physicians' utility, not only patients'.

where they should have been granted, introducing another type of gatekeeping failure as well as further losses in terms of expected health benefits.

Instead of a fee, the government could change the law, disallowing paid sick leave for patients with subjective diagnoses. Then the decision to be Strict or Lax would lie (at least formally) with the government, not with the physician. Again, it should be noted that a change of policy would not ease the inherent conflict between gatekeeping and healing; it would, however, remove the distributional part of the problem, which is related to the market structure.

Note, however, that this would create new incentives for corruption: in the case where physicians are Lax and subjective diagnoses qualify for paid sick leave, there is no reason for any patient to offer bribes to get a false sick leave, since she can get one for free. If that is no longer possible, there will be a potential gain of collusion between patient and physician, opening up the path for an illegal market for false sickness certificates.

Physicians' views: Empirical findings

The main result of the above analysis is that the gate is wide open: Unless physicians have rather extreme preferences for gatekeeping, they will, in practice, not function as gatekeepers at all. One implication is that it is, in essence, the worker who sicklists herself; the physician may give expert advice, but the final decision lies, in reality, with the patient. This conclusion may seem slightly absurd to an economist used to thinking of work simply as lost leisure; although Norwegian authorities do worry about high sicklisting rates, the vast majority of Norwegian workers are, after all, not sicklisted.

The above results were based on a highly stylized theoretical model, leaving open the possibility that crucial aspects of the problem have been overlooked. We have thus chosen to complement the theoretical analysis with in-depth interviews with Norwegian primary care physicians concerning their experiences with sickness certification.

The results from these interviews support our theoretical conclusion: The gate is open; sicklisting decisions are, to a large extent, left to patients.

Empirical method and procedures

During the fall of 2007 we sent 97 invitations to participate in focus group interviews to a random sample of coordinators for physicians' educational groups from the Norwegian counties of Hordaland and Oslo. Groups that responded by contacting us were included in the study consecutively. We continued interviewing groups until a point where no substantial new findings appeared; a point of data saturation (Glaser and Strauss 1967). In total we conducted five group interviews with a total of 28 physicians participating. The profile of the study sample was fairly similar to the population of Norwegian primary care physicians (see Table 1). All group participants knew each other from monthly meetings in the educational groups.

Table 1: Characteristics of the participating physicians and all primary care physicians in Norway

	Interview sample	Population of primary care physicians in Norway[*]
Physicians	28	3891
Percentage male	61%	68%
Mean age	45	48
Mean no of listed patients	1172	1196
Percentage with vacancies on list	38%	47%
Percentage specialists	46%	56%

[*] Source: NAV. Styringsdata for Fastlegeordningen, 4. kvartal 2007 [The Norwegian Labour and Welfare Administration. Data from the Regular General Practitioner Scheme 4. Quarter 2007] (http://www.nav.no/page?id=1073743257)

One of the authors[23] conducted the interviews following a semi-structured interview guide. The researcher attempted to promote free discussion around the main topics of the interview guide. These topics were based on earlier Scandinavian studies of primary care physicians' gatekeeper role in sickness certification (Carlsen and Norheim 2005 , Englund, Tibblin, and Svardsudd 2000, Gulbrandsen, Førde and Aasland 2002, Wahlström and Alexanderson 2004). Interviews lasted between 60 and 90 minutes and were recorded and transcribed word for word by a research assistant. The transcripts were then analysed using thematic content analysis (Green and Thorogood 2004) to extract shared concerns and fundamental premises.

The interview guide included the following themes:

- Experience with the gatekeeper role
- Perceptions of the regulations regarding sick leave and the concept of illness
- Trust and distrust in the patient-physician relationship
- Assessing patients' diagnoses and work capacity
- Examples of difficult decisions
- Suggestions for improvements of the system

The project was approved by the Privacy Ombudsman against the privacy and license requirements of the Data Inspectorate in relation to the Personal Data Act and Health Register Act. The participants signed an informed consent form containing information about their right to anonymity and right to

---

[23] Carlsen, who is trained as an anthropologist and experienced in conducting group interviews.

withdraw from the study at any point. The participants also filled in a brief questionnaire about themselves and their medical practice.

Some limitations of the interview method need to be mentioned: Firstly, in depth studies will normally be limited in terms of generalizability of empirical findings. The selected sample will for practical reasons usually be small and it will seldom be randomized and representative, thus the findings may not be valid for the majority of the study population (in our case: all Norwegian primary care physicians) or across contexts. These limitations apply also to our study. However, our study sample and the study population seem, according to the statitistics reported in Table 1, to be relatively comparable. In addition, our findings support previous international quantitative studies, which strengthens the probability that the findings are valid beyond the context of this study. More importantly, however, our motive for including qualitative data in the present study is not to derive statistically significant quantitative results, but rather to explore whether physicians' own experiences seem to be broadly consistent with our theoretical story. It is well acknowledged that qualitative findings may both serve to convey a deeper understanding of statistical facts as well as to build or support theory through tentative hypotheses (see eg. Huberman and Miles 2002). .

Group interviews are often thought to be inadequate for surveying the range of of different viewpoints within a group, but are better suited to expose generally accepted norms and attitudes (Kitzinger 1995). In the empirical part of this study, the aim was precisely to discover such mainstream attitudes about sickness certification among Norwegian primary care physicians. Hence, group discussion may be dominated by 'politically correct' statements, but this is not necessarily problematic. It is to be expected that assertive participants have some bearing on the statements and opinions of other participants, as they do in real life group settings. In the present study, participants generally appeared comfortable within the groups, and practically all participants contributed enthusiastically to the discussions, often challenging each others' opinions. We ascribe this frankness to the participants' familiarity with each other.

Below, we restrict ourselves to report findings which represent "consensus views", in the sense that many interviewees expressed similar views while no-one explicitly opposed these views. Individual statements not supported or not discussed by others will thus not be included.

Is the gate open?

The participating physicians consistently differentiated between short and long sickness spells. They defined short sick leaves as lasting up to two or three weeks. Usually, these spells were characterized as "unproblematic cases", usually related to bio-medically observable symptoms (i.e. verifiable diagnoses in our terminology). Longer spells or repeated spells were frequently perceived as more problematic, and were also more often characterized by diffuse symptoms, i.e. including a greater proportion of subjective diagnoses.

The physicians agreed that for shorter periods, they generally issued sickness certificates without questioning the patient's request at all. Such decisions were often made quickly as most patients had easy verifiable symptoms of common, un-dramatic ailments. Several participants said that patients asking for certification for a single, shorter period of sickness absence, for all practical purposes grant

themselves the certification. None of the physicians could come up with examples where they had denied patients shorter sick leaves.

You rarely question the patient's request for a sickness certification. It is hard to steer that process.

Physician A: For many conditions, the patient in principle grants himself or herself sick leave; we just sign. Physician B: It is like A. says: at the outset, they often sicklist themselves; I rarely deny a sick leave when they ask for it, at least.

The physicians felt that most people preferred working, and would therefore not attempt deceiving the physician, even though a minority might be trying to stay away from work from reasons other than strictly medical ones.

It is my opinion, and my point of departure, that people do want to work; hence I do not have the feeling that it is important to identify people who misuse the social security system.

When it comes to longer sickness spells, the physicians often found their role to be more problematic. The participants said they occasionally try to negotiate with the patient and make an effort to convince the patient of trying, sometimes little by little, to return to full time work; but even in these cases the physicians admitted to give in to the patient's request more often than not.

At the outset, they sicklist themselves, I rarely deny a sickness certificate when they ask for it. But I do play a role in getting them back. Then I negotiate and wring and wriggle and play a much more important role, I think.

We try to make them join us in a setup where they can function, where they feel that this will work out. Then you give and take a little: "Now you try out this, and then we'll talk in a week or two, and perhaps then you can manage to work even a bit more." If we just say straight no, the whole situation gets locked.

Relating to potential shirkers: Trust or distrust

In the interviews, some of the physicians claimed that they simply do not believe that patients attempt lying to their physician; firstly, because they believed few would want to shirk, and secondly, because they thought it would be difficult and uncomfortable for the patient to betray the reciprocal trust of the doctor – patient relationship. However, as discussions in the groups evolved, several of the participants reflected further on this issue or were opposed by colleagues in the group, and then acknowledged that it is logical that they probably sometimes are deceived by patients.

Physician C: If you arrive at the wrong diagnosis, then your conclusion about whether sick leave is required will of course be wrong too.
Physician D: Yes, obviously. I just expose my naïveté by saying that I don't discover much cheating.

Nevertheless, they also claimed that it is very difficult to reveal a lie indicating a subjective diagnosis:

It can be hard with musculoskeletal problems that you cannot really pin down. The patient may for example claim to feel pain, and you perform thorough tests, but cannot find anything, except that the

patient says he or she is in pain. That is difficult. To what extent should you say that you think this is only psychological?

Additionally they felt inhibited by a sense of duty to trust the patient and center the diagnosis on the patient's subjective experience; thus some said they had "decided to believe" the patients.

You cannot relate to a world in which everyone is a potential cheater – that you must stand there and hold back a whole nation of shirkers who really just want to be at home. Then you could not work. A horrible thought.

- or, it could simply be too socially distressing for the physician to relate to patients while thinking of them as liars:

[…] otherwise you have to sit and think that this person in front of me is lying, right to my face. And we cannot sit like that all day, thinking that way.

Another reason why the physicians felt hindered from revealing shirkers was the professional responsibility they have as the patient's regular primary care physician. This responsibility allegedly makes them averse to taking health risks on the patient's behalf, and incite primary care physicians to practice defensively.

Then there are those times when I've been really suspicious, but then I'm afraid that my own prejudices play a role. Am I certain about this? And if I think this guy is a real shirker, there's a good chance that I'm actually wrong; that this person seems unsympathetic to me does not necessarily mean that he does not qualify for a sick leave. For it has happened, some times, that I've been a bit harsh with the patient, and then they end up in a hospital policlinic and it turns out that "good heavens, man, this person has meniscrupture, I have to do an arthroscopy […]". And I think: Oh […], I've been rude to that patient.

I remember two rather young patients, 30 to 40 years, with substantial back pains, where I was skeptical to one of them. She has ended up being operated by specialists twice. Such experiences strengthen my conviction that I have to consider my own prejudices.

In conclusion, the physicians experienced the gatekeeper role regarding sickness certification as problematic, largely because they had reason to believe that a minority is granted sickness leave without medical reasons. However, they felt they had no choice but to trust their patients. Hence shorter sickness certifications were issued practically without questions asked. Longer sickness leaves were more often negotiated, and the physician would attempt to probe further for possibilities for the patient to return to work; nevertheless, these sick leaves were also usually granted. Hence, the physicians in our study regretted that they were, to a large extent, unable to function as gatekeepers for sickness certification.

I find this a difficult task. One feels a squeeze between what society wants and what the patient wants. Usually I feel that I work more for the patient than for society.

Summary and conclusions

It may be hard to imagine that a generous system of sick leave benefits could function reasonably well without effective gatekeeping. Yet our finding is that in a system like the Norwegian, with general care physicians as the main gatekeepers, physicians cannot be expected to fulfill that role. The views expressed by the general care physicians interviewed in our study confirm this: none of them could remember to have denied a short-term sick leave at request; and while they did sometimes, in cases of long-term sick leaves, try to persuade patients to return to work, a leave would usually be granted anyway if they were unable to convince the patient. Thus, the choice of whether or not to be sicklisted seems to lie, essentially, with the patient herself. These findings are in line with previous empirical work (e.g. Carlsen and Norheim 2005; Larsen et al 1994).

The intuitive reason for this lies in the asymmetric information characterizing the patient-physician relationship, and the existence of diagnoses justifying sick leave for which none of the symptoms can be verified by the physician. Thus, it will not be possible for a physician to distinguish a shirker claiming to experience the symptoms of a such diagnosis from a person who truly experiences these symtoms. If the physician tries to distinguish anyway, she will be bound to make mistakes. If a truly sick person is mistaken for a shirker and is thus not given appropriate treatment, including sick leave, the physician will, as a result, be a poorer healer. Hence, for patients with this type of diagnoses, there is an inherent conflict between the healer and the gatekeeper role.

Moreover, if physicians' income depends on the demand they face from patients, and patients are free to choose which physician to see, the market will amplify this gatekeeper – healer conflict. The reason is that a patient will typically not know, before seeing a physician, whether she does have symptoms that can be verified by the physician. She will thus avoid physicians with a reputation for being strict, even if she is among those patients who would, in fact, be equally well treated by a strict physician. The consequence is that strict physicians will not only be poorer healers for those with only subjective symptoms; they will face a substantially lower demand from all types of patients, leading to losses both in terms of income and the physician's success as a healer.

In Norway, the government has long been concerned about relatively high sickness absence rates. Considering that one cannot expect there to be effective gatekeeping at all, the puzzle is not really why sick leave rates are so high, but rather why they are so low. If it is patients themselves who make the decision of being present at work or being absent with full pay, and the large majority still choose to work, there must be substantial private benefits of being present at work. It is not hard to come up with examples of what kind of benefits this might be: non-pecuniary private benefits would include social interaction with co-workers or customers, social rewards when doing a good job, performing intrinsically interesting tasks, and the satisfaction of being important or useful to others (e.g. Brekke and Nyborg, 2008); pecuniary private benefits would include concerns about promotion and wage development over time (Markussen, 2009). Recent studies do indicate that sick leaves are higher for those with unpleasant working environments, low education level, or low pay (see, e.g., Ose 2005, Markussen et al. 2008); these are indeed groups that may be expected to have relatively low private benefits from being present at work.[24] If it is in fact the presence of these private benefits

---

[24] See e.g. Ose (2006), who finds that absenteeism is higher in noisy work environments. Our participants expressed related views, like the following: *The socially strongest often have interesting jobs with high responsibilities where they feel that*

that keep healthy workers from shirking, a different route to reduce sick leaves rates would be, rather than focusing on physicians' practice styles, to search for policies increasing the (relative) private benefits of working.

Our theoretical model was based on highly stylized assumptions, and relaxing these assumptions would of course modify the conclusions. Let us thus mention a few caveats. First, in the real world physicians differ, not only with respect to their strictness; with greater variation in physician characteristics, competition between physicians may be softened, which may reduce the market's amplification of the gatekeeper-healer conflict. However, the conflict will still be there. Secondly, in our model there were no costs of seeing a physician; in practice, patients do pay a modest fee, and in addition they have to spend some time and effort. Also, patients may experience discomfort when lying to their doctor. Taking these costs into account would modify our conclusions in the sense that the worker's preference to shirk would have to be stronger before she would choose seeing a doctor.

Third, and more importantly, we have disregarded that for some diagnoses, the prognosis for healing may be better if the patient is active, implying that a sick leave would *not* be recommended. For subjective diagnoses where this is the case, there is no conflict between the healer and gatekeeper role. Such concerns may possibly be one reason why the particpants in our study reported to negotiate with patients in order make them return to work.

In recent years, physicians appear to have strengthened their belief in the potential healing effects of patient activity for several diagnoses, such as back pain. If a particular subjective diagnosis which used to qualify for sick leave no longer does so, shirkers will quit claiming the symptoms of this diagnosis. As long as there exist some other subjective diagnoses requiring sick leave, shirkers can rather claim to experience those, and the problem will persist.

Hence, our main conclusion remains: Primary care physicians' role is basically that of being expert advisors to patients. The final decision of whether a sick leave should be granted, however, is in practice made by patients themselves. The relatively low absence rates in generous welfare states such as the Norwegian thus seem to imply that there must be substantial private benefits of being present at work, over and above contracted wages.

References

Alger, I., and C.A. Ma (2005): Moral Hazard, Insurance, and Some Collusion, *Journal of Economic Behavior and Organization* (50), 225-247.

Angell, M. (1993). The doctor as double agent. *Kennedy Inst Ethics J, 3*(3), 279-286.

Arnesen, T., & Fredriksen, S. (1995). Coping with obligations towards patient and society: an empirical study of attitudes and practice among Norwegian physicians. *Journal of Medical Ethics, 21*, 158-161.

---

they must, and want to, be present. Those who would like to sneak away are not the most resourceful ones.

Aronsson, Gunnar and Gustafsson, Klas (2005). Sickness Presenteeism: Prevalence, Attendance-Pressure Factors, and an Outline of a Model for Research. Journal of Occupational and Environmental Medicine, 47(9):958-966.

Ayres, P. J. (1996). Rationing health care: Views from general practice. *Social Science & Medicine, 42*(7), 1021-1025.

Brekke, K.A., and K. Nyborg (2008): Selfish Bakers, Caring Nurses? A Model of Work Motivation, *HERO Working Paper* 2008:1, Oslo: Health Economics Research Programme.

Carlsen, Benedicte, and Ole Frithjof Norheim (2003). Introduction of the patient-list system in general practice. Scandinavian Journal of Primary Health Care

Carlsen, B., & Norheim, O. F. (2005). "Saying no is no easy matter". A qualitative study of competing concerns in rationing decisions in general practice. *BMC Health Services Research, 5*(70).

Chalkley, M., and J.M. Malcolmson (1998): Contracting for Health Services when Patient Demand does not Reflect Quality, *Journal of Health Economics* 17, 1-20.

De Jaegher, K. and M. Jegers (2000): A model of physician behaviour with demand inducment, Journal of Health Economics, 19, 231-258.

Dranove, D. (1988): Demand Inducement and the Physician/Patient Relationship, *Economic Inquiry* 26, 251-298.

Eisenberg, J.M. (1986): *Doctors' Decisions and the Cost of Medical Care. The Reasons for Doctors' Practice Patterns and Ways to Change Them*, Ann Arbor, Michigan: Health Administration Press Perspectives.

Englund, L., Tibblin, G., & Svardsudd, K. (2000). Variations in sick-listing practice among male and female physicians of different specialities based on case vignettes. *Scandinavian Journal of Primary Health Care, 18*(1), 48-52.

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago: Aldine.

 Gulbrandsen, P., Førde, R., & Aasland, O. G. (2002). Hvordan har legen det som portvakt? *Tidsskrift for Den norske lægeforening, 122*, 1874-1879.

Huberman, A. M., and Matthew B. Miles (2002). The Qualitative Researcher's Companion: Classic and Contemporary Readings. Sage.

Kitzinger J: Qualitative Research: Introducing focus groups. *BMJ* 1995, 311:299-302.

Larsen, B., Forde, O., & Tellnes, G. (1994). Legens rolle i sykmelding [Physician's role in certification for sick-leave]. *Tidsskr Nor Laegeforen, 114*(12), 1442 - 1444.

Markussen, S. (2009): Home for free? The effect of sick-leave on earnings. Paper presented at Ragnar Frisch Centre research seminar, Oslo, 04.03.09.

Markussen, S., K. Røed, O.J. Røgeberg, S. Gaure (2008): The Anatomy of Absenteeism. Unpublished note, the Ragnar Frisch Centre for Economic Research, Oslo.

Medical Professionalism Project (2002): Medical Professionalism in the New Milennium: A Physician Charter. Project of the ABIM Foundation, ACP-ASIM Foundation, and European Federation of Internal Medicine. Annals of Internal Medicine 136 (3), 243-246.

Ose, S.O. (2005): Working conditions, compensation and absenteeism, *Journal of Health Economics* 24

Pearson, S. D. (2000). Caring and cost: the challenge for physician advocacy. *Ann Intern Med, 133*(2), 148-153.

Rochaix, L. (1989): Information Asymmetry and Search in the Market for Physician Services, *Journal of Health Economics* 8, 53-84.

Wahlström, R., & Alexanderson, K. (2004). Chapter 11. Physicians' sicklisting practices. *Scandinavian Journal of Public Health, 32*(0), 222-255. (1), 161-188.

World Medical Association (2006): World Medical Association International Code of Medical Ethics, http://www.wma.net/e/policy/c8.htm.