



Munich Personal RePEc Archive

Informal payments in developing countries' public health sector

Liu, Ting and Sun, Jiayin

11 October 2007

Online at <https://mpa.ub.uni-muenchen.de/5279/>

MPRA Paper No. 5279, posted 12 Oct 2007 UTC

Informal Payments in Developing Countries' Public Health Sectors*

Ting Liu[†] and Monic Jiayin Sun[‡]

September 2007

Abstract

In public health sectors of many developing countries, patients offer payments to their doctors outside the official payment channels. We argue that the fundamental cause of informal payments is that formal prices cannot fully differentiate patients' various needs. We compare welfare implications of different policies that can be used to regulate informal payments. Patient heterogeneity plays a central role in the comparison. Compared with banning informal payments, allowing them improves patient welfare if and only if patients' willingness to pay differs significantly. We also show that selling the right to choose physicians publicly always improves both patient welfare and social welfare.

JEL Codes: I11, I18, I38, H42

*We thank Albert Ma, Jacob Glazer, Bart Lipman, Dilip Mookherjee, and participants at Boston University Micro Theory Workshops for helpful suggestions.

[†]Boston University; tingl@bu.edu.

[‡]Boston University; monic@bu.edu.

1 Introduction

A World Bank report by Lewis (2000) begins with:

Informal payments in the health sector in Eastern Europe and Central Asia are emerging as a fundamental aspect of health care financing and a serious impediment to health care reform.

By definition, informal payments are those made to individuals or institutions in cash or in kind outside official channels for services that are meant to be covered by the public health care system. In China, for example, informal payments are often given in “red packets” in the public health sector and they have become a pressing social issue. The Chinese government treats such payments as bribes and has already imposed a national policy that whenever a doctor is found to accept informal payments, his license is immediately suspended by the Ministry of Health. Nevertheless, patients are still offering such payments. In 2004, Chinese doctors returned to patients or turned in to the state informal payments totalling 41.36 million RMB (roughly 5 million USD), as reported by the Ministry of Health. As there is little incentive for doctors to give up the informal payments, the actual amount of informal payments may be much higher than reported. Lewis (2000) lists the frequency of informal payments in some other countries in Table 1.

Many health care professionals believe that patients offer informal payments to induce more effort from the doctor, while others think the purpose is to conform with the social norm. As long as patients are rational economic agents, they must be paying for *something valuable*. It could be a higher level of doctor effort, the choice of a better doctor, or a better position on the waiting list. In any case, there must exist some mechanism which ensures that patients get better services when they pay more informally. For example, doctors are concerned about their reputation in a repeated game: if they do not react to informal payments in one period, they lose all future payments. Alternatively, they may simply feel guilty for not investing more effort when being paid more.

It is not our research objective to characterize this mechanism in a super-game. Instead, we take the mechanism as effective and simply assume that doctors react to informal payments. Essentially, we model informal payments as a device for patients to compete

TABLE 1 Estimated frequency of informal payments in selected countries from Lewis (2000). Armenia: non representative national sample data, inpatient care only. Poland: inpatient care only. Russian Federation: represents frequency of paying public hospitals but not at cash register.

Country	Year	Frequency of IP
Armenia	1999	91%
Vietnam	1992	81%
Azerbaijan	1995	78%
Poland	1998	78%
Kyrgyz Republic	1999	75%
Russian Federation	1997	74%

for better services. To fix ideas, we model better services as the option of seeing a more capable doctor, and implications on other dimensions of quality can be readily obtained from the same model.

We take the stance that patients have more information about doctors than the administrators. The Ministry of Health in China, for example, ranks doctors into different categories-experts, chief doctors, and ordinary doctors-and sets a uniform price for seeing doctors in each category. The ranking criteria include medical degree, years of practising, publications, and number of patients they have ever treated. Patients, on the other hand, may have a better judgement of a doctor's skill. They can gather information about the doctor from their own personal experience, their friends' recommendations, or even online reviews. Such firsthand information is crucial for doctors to build up a reputation among patients.

Patients, in turn, are willing to pay more to select doctors with a better reputation. This is the foundation of our model: the actual quality of care varies between different doctors who are paid the same through the formal channels. By modeling patients' competition through informal payments, we discuss policies that maximize patient welfare and social efficiency. Social efficiency does not depend on any transfers, and thus the amount of informal payments, while patient welfare depends crucially on such payments.

A crucial factor in welfare analysis is patient heterogeneity: informal payments should be allowed if and only if patients' willingness to pay is heterogeneous. Intuitively, allowing informal payments improves allocation efficiency, while banning them helps patients to save money. When patients differ greatly in their willingness to pay, achieving the optimal allocation is most important; if they differ little, the competition becomes wasteful as the allocation is barely better than random. We also analyze a second policy: publicly selling the right to choose doctors. We find that this policy can improve both patient welfare and social welfare compared with banning informal payments.

As we assume that patients' informal payments are not refunded even if they do not get to see the better doctor, our model is essentially an "all-pay auction". The analytical framework is conceptually similar to a "menu auction" used in Bernheim and Whinston (1986), Grossman and Helpman (1994). Bernheim and Whinston (1986) describe influence-seeking as an example of a "menu auction" game. In a menu auction, each of several principals who will be affected by an action offers a bid to an agent who will take that action. These bids take the form of schedules that associate a payment to the agent with each feasible option. Once the agent chooses an action, all of the principals pay the bids stipulated by their schedules. Bernheim and Whinston define an equilibrium in a menu auction as a set of contribution schedules such that each one is a best response to all of the others, and an action by the agent that maximizes her utility given the schedules that confront her. In our model, bids take the form of a simple one-dimensional offer rather than a schedule. Riley and Hillman (1989) study political rents and transfers in an all-pay auction similar to ours.

The literature on informal payments is quite limited. Lewis (2000) points out that informal payments arise to alleviate the mismatch between specialties needed and specialties provided. Garcia-Prado (2005) considers the severity of doctor punishment and the bargaining structure between patients and doctors in determining the equilibrium amount of informal payments. She does not model competition among patients. Biglaiser and Ma (2003) and Gonzalez (2004) study “moonlighting”, a related phenomenon in which public sector doctors work part time for private hospitals. They focus on how doctors divide their labor supply between the public and private sectors, in which reimbursement schemes are different.

Informal payments are, essentially, a form of corruption.¹ The focus of the corruption literature is often on strategic actions of the bureaucrats collecting bribes (Acemoglu and Verdier 2000, 1998, Lui 1985, Shleifer and Vishny 1993, 1992). We model passive doctors who simply treat the patient who pays the most and look for optimal regulatory policies that the social planner could employ.

Section 2 introduces the model. Section 3 then compares allowing and banning informal payments. Section 4 discusses the policy of publicly selling the right to choose doctors. Section 5 summarizes the welfare analysis. Section 6 discusses patients’ income constraints and concludes.

2 The Model

There are two patients and two doctors. Each doctor can treat only one patient. One patient’s illness is serious (H) and the other’s is common (L). One doctor is more capable and has a good reputation among patients (G); the other doctor is ordinary (B). Whether each doctor is more capable or mediocre is known to the patients but not to the social planner. The formal price is therefore the same for a patient seeing any of the two doctors and is normalized to zero.

If patient $i \in \{H, L\}$ is treated by doctor $j \in \{G, B\}$, his utility is v_i^j . Both patients want to be treated by the more capable doctor: $v_i^G > v_i^B$ for $i \in \{H, L\}$. In addition, the

¹See Bardhand (1997) for a review of the corruption literature.

seriously ill patient has a larger increase in utility when he is treated by the good doctor instead of the ordinary doctor:

$$v_H^G - v_H^B > v_L^G - v_L^B, \text{ or } \Delta H > \Delta L, \quad (1)$$

where $\Delta H = v_H^G - v_H^B$ is the seriously ill patient's (incremental) willingness to pay for being treated by the more capable doctor, and $\Delta L = v_L^G - v_L^B$ is the common patient's. Assumption (1) is a form of the single crossing property: the sicker is the patient, the more he gains from being treated by a more capable doctor.

The cost of treating a patient is also normalized to zero. Although we assume that the treatment cost does not vary across patients and doctors, it is straightforward to incorporate a more general cost structure. We use the zero treatment cost assumption and focus on patients' competition. Both doctors commit to selecting the patient who offers more informal payment. The two patients offer informal payments to attract the good doctor and neither of them offers any informal payment to the general doctor. There are two tie-breaking rules. First, when there is a tie in the two offers of informal payments, the more capable doctor randomly selects a patient. Second, a patient does not offer any informal payment when he is indifferent.

The first best allocation is that the seriously ill patient sees the more capable doctor. Patient welfare in this case is $v_H^G + v_L^B$. If there is a free market of health care, any price in the range $[\Delta L, \Delta H)$ sustains a Walrasian equilibrium in which the more capable doctor treats the seriously ill patient. The equilibrium in which price equals ΔL is associated with the highest level of patient welfare. Theoretically, a social planner could use a "first-price auction" to achieve the same level of patient welfare. However, the implementation costs of such a mechanism would be quite high.

3 When Should Informal Payments Be Banned?

Accepting informal payments is illegal in many countries. Now we discuss patient welfare when the social planner can successfully ban informal payments by some methods. In this

case, the allocation is random, resulting in patient welfare

$$\frac{1}{2}(v_H^G + v_H^B) + \frac{1}{2}(v_L^G + v_L^B).$$

Obviously, patient welfare is lower than in the first best allocation.

What happens if the social planner allows informal payments? Patients compete for the more capable doctor by offering informal payments. We analyze the following game.

Stage 1 Patients simultaneously offer informal payments, P_i , to the more capable doctor before diagnoses. Once a patient pays informal payments, the money cannot be refunded.

Stage 2 The more capable doctor commits to treating the patient who offers more informal payments. When both patients offer the same informal payments, the more capable doctor randomly select one patient to treat.

The motivation for using such a game is the following. First, both patients have to pay no matter who gets to see the more capable doctor. Since informal payments are under-the-table transactions, patients do not have a receipt for paying them. Consequently, once a patient delivers a "red pocket", the money cannot be refunded. On the other hand, even if a patient indeed has some proof of previous informal payments, he is unlikely to confront the doctor with request of being refunded. He may need to see the doctor again or his colleagues in the future and would rather confine to the social norm of being silent.

Second, doctors are fully rational and pick whichever patient that pays more. Our model can be thought of as a reduced form of a model in which the more capable doctor concerns for future profits. By committing to treat the patient who pays most, the doctor gives future patients a strong incentive to raise informal payments. Furthermore, the doctor cannot select a patient contingent on the severity of the patient's problem as informal payments are often paid before a diagnosis.

We look for Nash Equilibria of this game.

Proposition 1. *There is no pure strategy Nash Equilibrium.*

Proof. Given the patients' willingness to pay, $P_H \leq \Delta H$ and $P_L \leq \Delta L$. Suppose there is a pure strategy equilibrium (P_L^*, P_H^*) . First, suppose $P_L^* = P_H^*$. If the seriously ill patient deviates to offer $P_H^* + \epsilon$, he gets the more capable doctor for sure and suffers a payment loss of ϵ . As long as $\epsilon < \frac{1}{2}\Delta H$, he gets more utility. Second, suppose $0 < P_L^* < P_H^*$. The common patient benefits from deviating to offer zero informal payment. Third, suppose $0 = P_L^* < P_H^*$. The seriously ill patient benefits from deviating to offer $P_L^* + \epsilon$, as long as $\epsilon' < P_H^* - P_L^*$. Fourth, suppose $P_L^* > P_H^*$. The seriously ill patient's utility is $v_H^B - P_H$. He benefits from deviating to offer $P_L^* + \epsilon''$ as long as $\epsilon'' < \Delta H + P_H^* - P_L^*$. Summarizing the four cases, we concluded that there does not exist a pure strategy Nash Equilibrium. \square

This result comes from the continuity in patients' offers of informal payments. Each patient wants to outbid the other by only an infinitesimal amount and hence no pure strategy equilibrium can be sustained. We now turn to mixed strategy equilibria.

Let $F_i(x)$, with $i \in \{L, H\}$, denote patient i 's cumulative distribution function of offering informal payments.

Proposition 2. *The unique mixed strategy Nash Equilibrium is:*

$$F_L(x) = \begin{cases} 1 - \frac{\Delta L}{\Delta H} + \frac{x}{\Delta H}, & 0 \leq x \leq \Delta L, \\ 1, & x > \Delta L; \end{cases} \quad F_H(y) = \begin{cases} \frac{y}{\Delta L}, & 0 < y \leq \Delta L, \\ 1, & y > \Delta L. \end{cases}$$

Proof. When the seriously ill patient offers $x \in (0, \Delta L]$, his utility is

$$F_L(x)v_H^G + (1 - F_L(x))v_H^B - x = (1 - \frac{\Delta L}{\Delta H} + \frac{x}{\Delta H})v_H^G + (\frac{\Delta L}{\Delta H} - \frac{x}{\Delta H})v_H^B - x = v_H^G - \Delta L.$$

When he offers $x = 0$, his utility is

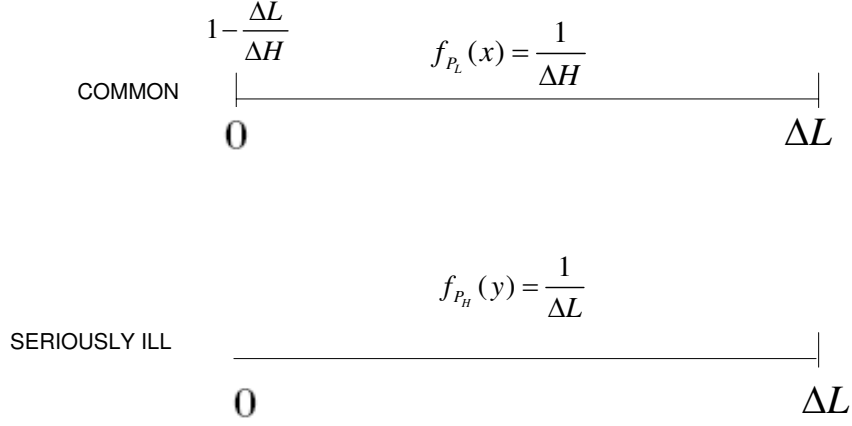
$$\frac{1}{2}(1 - \frac{\Delta L}{\Delta H})v_H^G + (\frac{1}{2} + \frac{\Delta L}{2\Delta H})v_H^B = \frac{1}{2}(v_H^G + v_H^B - \Delta L) < v_H^G - \Delta L.$$

When he offers $x > \Delta L$, his utility is $v_H^G - x < v_H^G - \Delta L$. Therefore the seriously ill patient's strategy in Proposition 2 is a best response to the common patient's strategy.

Similarly, when the common patient offers $y \in [0, \Delta L]$, his utility is

$$F_H(y)v_L^G + (1 - F_H(y))v_L^B - y = \frac{y}{\Delta L}v_L^G + (1 - \frac{y}{\Delta L})v_L^B - y = v_L^B.$$

FIGURE 1
 MIXED STRATEGY NASH EQUILIBRIUM WHEN INFORMAL PAYMENTS ARE ALLOWED



When he offers $y > \Delta L$, his utility is $v_L^G - y < v_L^B$. Therefore the common patient is also playing a best response. Proof of uniqueness of the equilibrium is left in the Appendix. \square

Figure 1 illustrates the two cumulative distribution functions in Proposition 2.

The seriously ill patient's offer is uniformly distributed in $(0, \Delta L]$ with density $\frac{1}{\Delta L}$. Notice that offering zero informal payment is not the seriously ill patient's best response against the common patient's strategy. This is reflected in Figure 2 as the P_H line has an open left support. The common patient offers zero informal payment with probability $1 - \frac{\Delta L}{\Delta H}$. He makes an offer in $[0, \Delta L]$ under the uniform density $\frac{1}{\Delta H}$. He never pays more than ΔL .

Proposition 2 has four implications. First, the lower is $\frac{\Delta L}{\Delta H}$, the more likely it is that the common patient offers zero informal payment. When the seriously ill patient is willing to pay a great deal more for the more capable doctor than the common patient is, the seriously ill patient offers a bulky "red packet" of informal payment. The common patient has little hope to win the competition, and meanwhile his incremental utility from being treated by the more capable doctor is low. As a result, he would rather

quit the competition and save some money. On the other hand, when two patients are willing to pay exactly the same for the more capable doctor, they both offer a strictly positive amount of informal payment. In this case, their random offers turn into a wasteful competition as the allocation is such that each patient gets the more capable doctor with probability .5. If the two patients can both commit to not paying informal payments, both are better off.

Second, the seriously ill patient is more likely to be treated by the more capable doctor. The probability of the first best allocation is

$$\Pr(P_H > P_L) = 1 - \frac{\Delta L}{\Delta H} + \int_0^{\Delta L} \left(\int_{P_L}^{\Delta L} \frac{1}{\Delta L} dy \right) \frac{1}{\Delta H} dx = 1 - \frac{1}{2} \frac{\Delta L}{\Delta H} > \frac{1}{2}$$

The stronger the heterogeneity in the patients' willingness to pay, the higher the probability for the seriously ill patient to be treated by the more capable doctor.

Third, the ratio of the two patients' expected informal payments equals to the ratio of their willingness to pay. The expected value of informal payments offered by the common patient is

$$E(P_L) = \int_0^{\Delta L} \frac{x}{\Delta H} dx = \frac{\Delta L^2}{2\Delta H},$$

and by the seriously ill patient is

$$E(P_H) = \int_0^{\Delta L} \frac{x}{\Delta L} dx = \frac{\Delta L}{2}.$$

Therefore, $\frac{E(P_L)}{E(P_H)} = \frac{\Delta L}{\Delta H}$. In other words, if the seriously ill patient's willingness to pay is two times the common patient's, his expected informal payment is also two times the common patient's.

Fourth, the total amount of informal payments, $E(P_L) + E(P_H) = \frac{\Delta L^2}{2\Delta H} + \frac{\Delta L}{2}$, increases in the common patient's willingness to pay and decreases in the seriously ill patient's. As the common patient pays more, the seriously ill patient also pays more, and the total amount of informal payments increases. On the other hand, when the seriously ill patient pays more, the common patient is more likely to quit the competition, which reduces the total amount of informal payments.

When informal payments are allowed, the expected utility of the common patient is v_L^B and that of the seriously ill patient is $v_H^G - \Delta L$. Compared with the case in which informal payments are banned, the common patient is always worse off. The seriously ill patient is also worse off when $\Delta H < 2\Delta L$. As long as $\Delta H < 3\Delta L$, allowing informal payments decreases aggregate patient welfare.

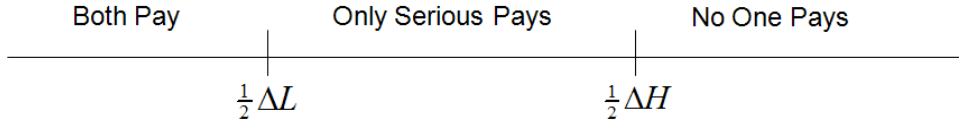
Nevertheless, the seriously ill patient is always more likely to see the more capable doctor when informal payments are allowed, and as a consequence social welfare is improved. The welfare comparison justifies some developing countries' ban of informal payments: when the social planner's goal is to maximize patient welfare, he should ban informal payments whenever patients do not differ much in their willingness to pay for the more capable doctors.

4 Selling the Right to Choose Doctors

Banning informal payments often involves high monitoring costs and does not always improve patient welfare. In this section, we examine an alternative policy that the social planner can resort to: publicly selling the right to choose doctors. To be precise, the social planner can set a non-refundable price p and make sure that whoever pays this price gets the right to choose doctors. If both or neither patients pay, doctors are allocated randomly. Assume that a patient does not pay p when he is indifferent between paying the price and not paying it. To figure out the price that maximizes patient welfare, we first show how patients react to different prices.

Proposition 3. *The Nash Equilibrium is that both patients pay if $p < \frac{1}{2}\Delta L$, only the seriously ill patient pays if $\frac{1}{2}\Delta L \leq p < \frac{1}{2}\Delta H$ and no patient pays if $p \geq \frac{1}{2}\Delta H$.*

Proof. We characterize ranges of p that sustain each type of equilibrium. Start with the “both pay” equilibrium. For this equilibrium to hold, $\frac{1}{2}v_L^G + \frac{1}{2}v_L^B - p > v_L^B$ for the common patient and $\frac{1}{2}v_H^G + \frac{1}{2}v_H^B - p > v_H^G$ for the seriously ill patient. Therefore, this equilibrium is sustained if $p < \frac{1}{2}\Delta L$. Similarly, the “only seriously ill patient pays” equilibrium is sustained if the common patient faces $p \geq \frac{1}{2}\Delta L$ and the seriously ill patient faces

FIGURE 2**NASH EQUILIBRIUM WHEN SELLING THE RIGHT TO CHOOSE DOCTORS**

$p < \frac{1}{2}\Delta H$. These two conditions combine to $\frac{1}{2}\Delta L \leq p < \frac{1}{2}\Delta H$. If $p \geq \frac{1}{2}\Delta H$, each patient is better off not paying given that the other patient does not pay. No price can sustain an equilibrium in which only the common patient pays. \square

Figure 2 illustrate Proposition 3.

What price does the social planner choose to maximize patient welfare? First realize that the social planner never sets a positive price lower than $\frac{1}{2}\Delta L$. If he does, both patients pay, which leads to more out-of-pocket spending and no improvement in allocation efficiency. When the price is set in the range $[\frac{1}{2}\Delta L, \frac{1}{2}\Delta H)$, the efficient allocation is induced. Setting $p = \frac{1}{2}\Delta L$ brings higher patient welfare than any other prices in this range. When the price is greater than $\frac{1}{2}\Delta H$, no one pays and again allocation efficiency is not achieved, and hence the social planner never sets $p \geq \frac{1}{2}\Delta H$. Summarizing the three cases, the essential question is whether to set $p = \frac{1}{2}\Delta L$ or $p = 0$.

Proposition 4. *A social planner who maximizes patient welfare sets $p = \frac{1}{2}\Delta L$ if $\Delta H > 2\Delta L$, and $p = 0$ otherwise.*

Proof. Patient welfare is $v_L^B + v_H^G - \frac{1}{2}\Delta L$ when $p = \frac{1}{2}\Delta L$, and $\frac{1}{2}(v_L^B + v_L^G) + \frac{1}{2}(v_H^B + v_H^G)$ when $p = 0$. The former is higher if and only if $\Delta H > 2\Delta L$. \square

Intuitively, the gain from the allocation efficiency compensates the extra payments only when patients' willingness to pay is quite different. When the social planner sets $p = 0$, he is effectively banning informal payments. In the next section, we characterize conditions for selling the right to choose doctors to be strictly superior to banning informal payments.

TABLE 2
 $\frac{\Delta L}{\Delta H} < \frac{1}{2}$.

Welfare Comparison When Patients' Willingness to Pay is Different:

	Common patient	Seriously ill patient	Patient Welfare	Prob (Match)
Ban	$\frac{1}{2}(v_L^B + v_L^G)$	$\frac{1}{2}(v_H^G + v_H^B)$	$\frac{1}{2}(v_L^B + v_L^G) + \frac{1}{2}(v_H^G + v_H^B)$	$\frac{1}{2}$
Allow	v_L^B	$v_H^G - \Delta L$	$v_L^B + v_H^G - \Delta L$	$1 - \frac{\Delta L}{2\Delta H}$
Sell	v_L^B	$v_H^G - \frac{1}{2}\Delta L$	$v_L^B + v_H^G - \frac{1}{2}\Delta L$	1

5 Welfare Analysis

We compare patient and social welfare in different policy scenarios, and discuss the social planner's choice of the optimal policy. We consider two cases. When $0 \leq \frac{\Delta L}{\Delta H} < \frac{1}{2}$, we say that the two patients' willingness to pay is different; when $\frac{\Delta L}{\Delta H} \geq \frac{1}{2}$, we say that the two patients' willingness to pay is similar.

5.1 Patients' Willingness to Pay is Different

When $\frac{\Delta L}{\Delta H} < \frac{1}{2}$, we summarize each patient's utility, aggregate patient welfare and the probability of achieving the efficient match in Table 2. As informal payments are transfers between patients and doctors, the probability of achieving the efficient match is proportional to the level of social welfare.

Table 2 has several notable features. First, the common patient's welfare is maximized when informal payments are banned. This captures the intuition that when a patient's illness is not serious, he is not willing to pay informally to guarantee treatment from a doctor with a better reputation. Therefore, he prefers the social planner to ban informal payments altogether.

Second, allowing informal payments improves social welfare more than banning them:

the more capable doctor is allocated to the seriously ill patient with a higher probability. However, allowing informal payments does not always improve patient welfare. On one hand, there is a higher probability of achieving efficient allocation. On the other hand, patients have to pay more out of their pockets. When $\frac{1}{3} \leq \frac{\Delta L}{\Delta H} < \frac{1}{2}$, allowing informal payments lowers patient welfare. When $0 \leq \frac{\Delta L}{\Delta H} < \frac{1}{3}$, allowing informal payments improves patient welfare.

Third, publicly selling the right to choose doctors is always a superior policy: it maximizes both patient and social welfare. Under this policy scheme, the common patient never pays, which enables the seriously ill patient to see the more capable doctor at a lower cost.

5.2 Patients' Willingness to Pay is Similar

When $\frac{\Delta L}{\Delta H} \geq \frac{1}{2}$, a social planner who maximizes patient welfare sets the price to zero when selling the right to choose doctors. In other words, he is banning informal payments, which improves patient welfare more than allowing them. Allowing informal payments, however, corresponds to a higher probability of achieving efficient allocation, and thus a higher social welfare.

5.3 The Best Policy for the Social Planner

A social planner who maximizes patient welfare chooses among no regulation (allow), banning informal payments (ban) and selling the right to choose doctors (sell). Table 3 summarizes the ranking of patient welfare in the three regimes.

We highlight two results in Table 3. First, selling the right to choose doctors is always the best policy. The feasibility of this policy depends on the extent to which a social planner can learn the patients' willingness to pay. Surveys and other forms of research on patient-doctor relationships can be useful.

Second, as the difference in patients' willingness to pay becomes larger, allowing informal payments starts to dominate banning them. A large gain in the allocation efficiency justifies the possible waste of informal payments.

TABLE 3

Ranking of Different Policies By Patient Welfare

Patient Heterogeneity	Condition	Patient Welfare
High	$0 \leq \frac{\Delta L}{\Delta H} < \frac{1}{3}$	Ban \prec Allow \prec Sell
Medium	$\frac{1}{3} \leq \frac{\Delta L}{\Delta H} < \frac{1}{2}$	Allow \prec Ban \prec Sell
Low	$\frac{1}{2} \leq \frac{\Delta L}{\Delta H} < 1$	Allow \prec Ban \equiv Sell

6 Discussion and Conclusion

6.1 Income Constraints

We have assumed that both patients can offer as much informal payment as they want. In reality, patients may have income constraints. In particular, patients with serious problems may not be able to offer enough informal payments to attract the more capable doctor. Suppose the seriously ill patient's income, I_H , is less than ΔL ; his income constraint binds.

Consider the game in Section 3 again when informal payments are allowed. As before, there is no pure strategy Nash Equilibrium. The unique mixed strategy Nash Equilibrium becomes:

$$F_L(x) = \begin{cases} 1 - \frac{I_H}{\Delta H} + \frac{x}{\Delta H}, & 0 \leq x \leq I_H; \\ 1, & x > I_H. \end{cases} \quad F_H(y) = \begin{cases} 1 - \frac{I_H}{\Delta L} + \frac{y}{\Delta L}, & 0 \leq y \leq I_H; \\ 1, & y > I_H. \end{cases}$$

Both patients now have a positive probability of offering no informal payment. The probability of achieving the first best allocation becomes

$$Pr(P_H > P_L) = \left(1 - \frac{I_H}{\Delta H}\right) \frac{I_H}{\Delta L} + \int_0^{I_H} \frac{I_H - P_L}{\Delta L} \frac{1}{\Delta H} dx$$

$$= \frac{I_H}{\Delta L} \left(1 - \frac{I_H}{2\Delta H}\right) \begin{cases} < \frac{1}{2}, & \text{if } 0 \leq I_H < \Delta H - \sqrt{\Delta H(\Delta H - \Delta L)}; \\ \geq \frac{1}{2}, & \text{if } \Delta H - \sqrt{\Delta H(\Delta H - \Delta L)} \leq I_H \leq \Delta L. \end{cases}$$

Recall that with no income constraints, social welfare is higher if informal payments are allowed than if they are banned. With income constraints, when I_H is small, which means that the seriously ill patient is poor, informal payments do not always improve efficiency. Banning informal payments in this case improves both patient and social welfare. When the seriously ill patient is not too poor, allowing informal payments has similar consequences as before: a gain in efficiency and a loss in patients' wealth. In general, for a seriously ill patient, having an income constraint makes allowing informal payments less attractive than banning them.

6.2 Conclusion

Informal payments can be attributed to complicated reasons in reality. It is not our purpose to analyze every possible reason; rather, we take the reasons as given and analyze policies that can potentially improve patient welfare.

Our analysis has several implications. First, whether the social planner should allow informal payments depends crucially on patient heterogeneity. Banning informal payments is not always the right choice, nor is allowing them. When patients are very heterogenous, informal payments can work to improve patient welfare. When patients are more or less the same, informal payments become a waste of patients' money.

Second, publicly selling the right to choose doctors may alleviate the problem of wasteful competition. Some analysts in China have already proposed this policy² and our analysis reveals that it is worth trying. Patients can pay for the right to choose better doctors, as well as more advanced medical procedures or services.

Third, our model yields additional implications when one assumes that the variation in patients' willingness to pay comes from their wealth. First of all, the wealthy patient is more likely to be treated by the more capable doctor. Informal payments give the the wealthy patient a competitive edge over the poor one. The more different the patients'

²<http://news.qq.com/a/20050114/000346.htm>, in Chinese, accessed December 2006.

levels of wealth, the more likely that the wealthy patient sees the more capable doctor. Second, as the poor patient's wealth approaches zero, the probability that the wealthy patient gets to see the more capable doctor approaches one.

Fourth, privatizing the public health sector, as proposed by some policy analysts, may not be a good idea. The analysts argue that in a free market of health care, price would efficiently allocate resources and social welfare is maximized. We agree with this argument but pay more attention to patient welfare, which may shrink severely in a free health care market. This helps to explain why few countries adopt a purely private health care system. Essentially, doctors may have strong bargaining power over their patients and, if so, when the more capable doctor were to set the price, he would make it as high as possible. Whether patient welfare can be improved by privatization depends again on the tradeoff between improvement in the allocation efficiency and the loss from payments.

Besides, our model suggests that privatization leads to lower patient welfare than selling the right to choose doctors in the public system. The first best allocation is achieved in both regimes, but the price of the more capable doctor is lower in the latter.

Last, one popular view is that informal payments result from a low level of doctors' wages in the public health sector. Consequently, raising the average doctor's wage is proposed to eliminate informal payments. We disagree with this proposal. An important reason for informal payments is the social planner's lack of information. As long as doctors' wages do not fully incorporate patients' information, informal payments will not completely disappear.

For future research, we are interested in welfare consequences of relevant policies when public and private hospitals co-exist, when some physicians are altruistic, or when doctors actively compete to get informal payments.

Appendix

Complete Proof of Proposition 2. A mixed strategy Nash Equilibrium is characterized by cumulative density functions of the two patients' offers, $F_L(x)$ and $F_H(y)$, and the supports, $[\underline{P}_L, \overline{P}_L]$ and $[\underline{P}_H, \overline{P}_H]$. We prove Proposition 2 in six steps as follows.

Step 1: the upper bounds of the two patients' offers are the same: $\overline{P}_L = \overline{P}_H = \overline{P}$.

When patient i 's offer is strictly bigger than the highest possible offer of the other patient, patient i could benefit from deviating to a smaller offer. As long as the new offer is still higher than the highest possible offer of the other patient, patient i still guarantees treatment from the more capable doctor.

Step 2: the upper bound is smaller than ΔL : $\overline{P} \leq \Delta L$. The common patient can always pay nothing and obtain his reservation utility v_L^B . If he offers more than ΔL , his utility is lower than v_L^B .

Step 3: the lower bounds of the two patients' offers are both zero. If the lower bound of patient i 's offer, \underline{P}_i , is strictly positive. The other patient, j , would not offer any amount in $(0, \underline{P}_i)$, as $P_j = 0$ strictly dominates any offer in the interval. Given this, patient i can profitably deviate to offer $\underline{P}'_i = \underline{P}_i - \epsilon$, where $\epsilon > 0$ is a small number.

Step 4: both distributions of offers are continuous. Suppose patient i makes an offer $P_i \in (0, \overline{P}]$ with probability $q > 0$ in equilibrium. He then makes an offer in the interval $(P_i - \epsilon, P_i)$ with zero probability, where $\epsilon > 0$ is a small number. Therefore, patient j also makes an offer in the interval $(P_i - \epsilon, P_i)$ with zero probability, as any such offer is strictly dominated by $P_j = P_i - \epsilon$. Given patient j 's strategy, patient i has a profitable deviation to $P'_i = P_i - \frac{\epsilon}{2}$.

Step 5: any mixed strategy equilibrium is characterized by
$$\begin{cases} F_L(x) = \frac{\Delta H - \overline{P} + x}{\Delta H}; \\ F_H(y) = \frac{\Delta L - \overline{P} + y}{\Delta L}. \end{cases}$$

In any mixed strategy Nash Equilibrium, each patient must be indifferent across his offers.

Therefore,

$$(v_i^G - P_i) \Pr(P_i > P_j) + (v_i^B - P_i) \Pr(P_i < P_j) = U_i,$$

where U_i is patient i 's equilibrium level of utility. As a result, $F_{P_j}(x) = \frac{x - v_i^B + U_i}{\Delta_i}$. Now impose the conditions $F_L(\bar{P}_L) = 1$ and $F_H(\bar{P}_H) = 1$, we get

$$U_H = \Delta H + v_H^B - \bar{P}, \quad U_L = \Delta L + v_L^B - \bar{P}.$$

Therefore, the two distribution functions are
$$\begin{cases} F_H(x) = \frac{x + \Delta L - \bar{P}}{\Delta L}; \\ F_L(y) = \frac{y + \Delta H - \bar{P}}{\Delta H}. \end{cases}$$

Step 6: at most one patient offers zero informal payment with positive probability. Suppose both patients offer zero informal payment with positive probability. Patient i can then profitably deviate to putting that positive probability to $\epsilon > 0$ instead of zero, where ϵ is a small number.

Step 7: the upper bound $\bar{P} = \Delta L$. From step 4, we know that $F_L(0) = \frac{\Delta H - \bar{P}}{\Delta H}$. Since $\bar{P} < \Delta H$, it must be that $F_L(0) > 0$. From step 5, $F_H(0)$ must be zero, which implies that $\bar{P} = \Delta L$. □

References

- ACEMOGLU, D., AND T. VERDIER (1998): “Property Rights, Corruption and the Location of Talent: a General Equilibrium Approach,” *Economic Journal*, Vol. 108(No. 450), 1381–1403.
- (2000): “The Choice between Market Failures and Corruption,” *American Economics Review*, Vol. 90(No. 1), 194–211.
- BARDHAND, P. (1997): “Corruption and Development: A Review of Issues,” *Journal of Economics Literature*, Vol. 35(No. 3), 1320–1346.
- BERNHEIM, B. D., AND M. WHINSTON (1986): “Menu Auctions, Resource Allocation, and Economic Influence,” *Quarterly Journal of Economics*, 101, 1–31.
- BIGLAISER, G., AND A. C. MA (2003): “Moonlighting: Public Service and Private Practice,” *Working Paper*.
- GARCIA-PRADO, A. (2005): “An Analysis of Under-reporting and Informal Payments among Physicians,” *Unpublished working paper*, Vol. 8(No. 5), 67–98.
- GONZALEZ, P. (2004): “Should Physicians’ Dual Practice be Limited? An Incentive Approach,” *Health Economics*, 13, 505–524.
- GROSSMAN, G., AND E. HELPMAN (1994): “Protection for Sale,” *American Economic Review*, 84, 833–850.
- LEWIS, M. (2000): “Who is Paying for Health Care in Europe and Central China?,” *The World Bank*.
- LUI, F. T. (1985): “An Equilibrium Queueing Model of Bribery,” *Journal of Political Economy*, Vol. 93(No. 4), 760–781.
- RILEY, J., AND A. HILLMAN (1989): “Politically Contestable Rents and Transfers,” *Economics and Politics*, 1, 17–39.
- SHLEIFER, A., AND R. VISHNY (1992): “Pervasive Shortages under Socialism,” *RAND Journal of Economics*, Vol. 23(No. 2), 237–246.
- (1993): “Corruption,” *Quarterly Journal of Economics*, Vol. 108(No. 3), 599–617.