AUSTRALIAN
ECONOMIC PAPERS

PRICE DISCRIMINATION IN PUBLIC HEALTHCARE*

NATHAN BERG
University of Otago

JEONG-YOO KIM
Kyung Hee University

Could a public healthcare system use price discrimination—paying medical service providers different fees, depending on the service provider's quality—lead to improvements in social welfare? We show that differentiating medical fees by quality increases social welfare relative to uniform pricing (i.e. quality-invariant fee schedules) whenever hospitals and doctors have private information about their own ability. We also show that by moving from uniform to differentiated medical fees, the public healthcare system can effectively incentivise good doctors and hospitals (i.e. low-cost-types) to provide even higher levels of quality than they would under complete information. In the socially optimal quality-differentiated medical fee system, low-cost-type medical-service providers enjoy a rent due to their informational advantage. Informational rent is socially beneficial because it gives service providers a strong incentive to invest in the extra training required to deliver high-quality services at low cost, providing yet another efficiency gain from quality-differentiated medical fees.

JEL code: I13.

I. Introduction

This paper has six main theoretical findings regarding public healthcare systems that use quality-differentiated medical fees (i.e. paying more to hospitals and doctors that are able to provide better clinical outcomes at lower cost) in environments where quality or cost types are private information, known only to the service provider about itself. (i) Low-cost hospitals provide the first-best socially optimal quality of medical service that would be provided under complete information. (ii) High-cost hospitals do worse than they would under complete information. (iii) Low-cost hospitals provide higher-quality service than high-cost types do. (iv) Low-cost hospitals receive higher medical fees than high-cost hospitals. (v) The quality-differentiated medical fee

doi: 10.1111/1467-8454.12114
Correspondence: Nathan Berg, Department of Economics, University of Otago, PO Box 56, Dunedin 9054, New Zealand. nathan.berg@otago.ac.nz.
(Accepted for publication 30 December 2017)
*Funding support for this research is gratefully acknowledged from the following sources: Australian Research Council (ARC), award number DP150100242; a grant from Kyung Hee University in 2016, KHU-20160685; and a University of Otago Research Grant 2018.

structure (under the assumptions given in our model) increases social welfare relative to uniform pricing (i.e. quality-invariant medical fee schedules) whenever hospitals and doctors have private information about their own ability. And (vi), the optimal quality-differentiated medical fee structure effectively incentivises good doctors and hospitals (i.e. low-cost types) to provide higher levels of quality than they would under complete information. These findings are consistent with the standard results of 'no distortion at the top' and 'some distortion at the bottom' wherein low-cost types provide higher quality (and charge more) than high-cost types do. Our findings hold whether the number of types of quality among medical service providers is finite or a continuum.

In many countries, the public healthcare system pays medical service providers according to a uniform fee schedule, that is, invariant with respect to quality. In recognition of inefficiencies caused by fee systems that provide little reward for high-quality medical service, many developed countries, such as France, Australia and the United States, are adopting new legislation and regulatory policies aimed at making medical fees contingent on quality to a greater extent. Japan's public healthcare system has already adopted the competition principle into determination of its medical fees, allowing more competent doctors to charge larger fees, especially for surgery. According to Japan's quality-contingent medical fee schedule, there is a limit in place stipulating that fees paid to the highest-quality surgeon can be, at most, slightly less than double the fee paid to the lowest-quality surgeon. The ministry of health and welfare (MHW) of Korea is also considering the introduction of differentiated medical fees.[1] For example, doctors with identical specialisations working in the same hospital might be allowed to charge different fees for the same treatment depending on previous patient outcomes or other measurable differences in their career histories,[2] whereas under current institutions governing public healthcare in Korea, a uniform medical fee is set by the government for any given medical service across different doctors and hospitals. Under Korea's current uniform fee system, the broadly defined ability of doctors (i.e. cost type or, equivalently, quality type) is not well-reflected in the medical fees they are paid. This gap implies that the incentive for doctors to improve their medical ability or otherwise choosing to supply high-quality service is too weak, thus leading to social inefficiency.

Uniform medical fees (as a state mandate) lead some patients to find grey-market channels for compensating doctors for higher-quality service with so-called 'gratitude fees'. Under-the-table payments to high-skill doctors, or highly sought after specialists, was commonplace (and apparently continues) in Hungary[3] (OECD, 2005), other Eastern European countries (Chawla *et al.,* 1998; Lewis, 2000; Miller & Koshechkina, 2000; Balabanova & McKee, 2002; Ensor, 2004; Allin *et al.,* 2006; Atanasova *et al.,* 2015), Korea, and elsewhere. The extent of gratitude

[1] Different versions of price discrimination (e.g. measuring quality, competence or ability levels of hospitals and doctors) are possible in the determination of medical fees paid by public healthcare systems. In Korea, some proposals stipulate that price discrimination would be based on the number of patient treatments. Price discrimination based on numbers of patients treated can, for example, be found in Taiwan. In 2012, however, Korea began considering a new form of price discrimination based on medical cost or (what is equivalent, given service providers with different production technologies) on productivity.

[2] This new policy is distinguished from the standard fee-for-service (FFS) payment system in which a single fee is charged for each unbundled and separate service, because it allows different fees to be charged based on different levels of expected quality for the very same service. However, if we interpret FFS as charging different fees for quality-differentiated versions of the same procedure or treatment (i.e. charging different fees for different qualities of service), then FFS can be viewed as the system of differentiated medical fees (contingent on quality) that we consider in this paper.

[3] Opinion polls and a referendum movement by doctors in Hungary to stop the practice of gratitude money reveal a broad range of interpretations: "Among doctors, 89% have a negative opinion of gratitude money, but 78% report that they accept these payments nevertheless. Both doctors and patients consider it a form of corruption." Accessed at http://www.politics.hu/20131016/doctors-may-call-referendum-on-gratitude-money/

payments, which appear across a diverse range of countries where uniform medical fees are imposed by the public healthcare system, is an important stylised fact that speaks to just how strong the two-sided efficiency gains from price differentiation for medical service providers can be.

The Affordable Care Act (ACA) recently implemented in the United States has already led to important changes in fees paid to hospitals and doctors depending on different dimensions of service providers' quality. For example, the ACA actively encourages participation in accountable care organisations (ACOs) while offering bonus payments for service providers that serve a sufficient number of medicare patients while achieving other patient-care benchmarks. Research has already appeared measuring substantial effects of legislative change (passage of ACA) and ongoing regulatory changes (e.g. interpretation of rules for participating in state healthcare exchanges) on rates of insurance coverage among US healthcare consumers (e.g. Courtemanche *et al.,* 2014). Relatively little research has, as of yet, looked at the extent to which ACA encourages new forms of price discrimination or the possibility that such price discrimination (enabled by new legislation and regulatory changes) might achieve substantial gains in social welfare by means of improved informational efficiency in the supply decisions of hospitals, doctors and other medical services providers with heterogeneous quality.

In this paper, we analyse the differentiation of medical fees payable to doctors of different types from an informational point of view. The public healthcare system can be defined by the organisation of medical institutions, consumers (population) and resources that deliver healthcare services to meet the health needs of the population. Thus, it includes not only the supply side of medical services, but also the population. Therefore, we can assume that the objective function of the public healthcare system is the sum of the consumer surplus and some (weighted) profits of the medical institutions. One serious problem when the public healthcare system designs the price system is, however, that it has limited information about the ability of individual physicians. Using the mechanism design approach, we characterise a socially optimal medical fee system when physicians are privately informed about their medical ability. Our analysis provides a rationale for public healthcare systems to allow differentiated medical fees to hospitals and doctors depending on quality. Our model demonstrates that payment of medical fees differentiated by quality of service turns out to be a socially efficient form of second-degree price discrimination.

Maskin and Riley (1984) provide seminal formal analysis of second-degree price discrimination. Their paper confirms the widespread wisdom that a monopolist can increase its profit by engaging in second-degree price discrimination (for which Maskin and Riley characterise the optimal pricing schedule).[4] Our paper is distinct from Maskin and Riley (1984) in two main respects. First, their model of price discrimination concerns consumers with private information about their own valuation of the product that a monopolist sells. In contrast, our model addresses price discrimination with respect to medical fees paid to hospitals and doctors that have private information about their own ability. Second, in Maskin and Riley's (1984) model, a profit-maximising monopolist price discriminates, whereas in our model a welfare-maximising government carries out price discrimination. Of course, many previous studies use models in which—similar to ours—a regulator seeks to maximise social welfare. In our model, however, it is the

[4] Following the pioneering work by Mussa and Rosen (1978), there is by now a vast literature on second-degree price discrimination (e.g. Phlips, 1983; Matthews & Moore, 1987; Srinagesh & Bradburd, 1989). This literature typically finds that the quality of a good is distorted only for consumers with high valuations. Srinagesh and Bradburd (1989), however, provide a novel argument implying that quality is distorted only for consumers with low valuations of the good. See Stole (2007) for a survey on price discrimination.

level of hospitals and doctors' ability, that is, private information rather than consumer characteristics, which are private information (about which the regulator is uninformed) in previous studies. For example, Goldman *et al*. (1984) consider social-welfare-maximising nonuniform prices when consumers' willingness to pay is distributed over some closed interval, which is private information. Mirman and Sibley (1980) extend this problem to the case of a monopoly that produces multiple products. And Sharkey and Sibley (1993) consider a similar problem allowing for variable weights assigned to each of the two consumer types. Our problem draws motivation from current events which reveal significant numbers of legislative and regulatory initiatives around the world (as mentioned above) aimed at improving social efficiency by introducing price discrimination into the health sector, where public healthcare systems are the largest buyers of healthcare services.

Reinhardt (2006) describes second-degree price discrimination in the United States medical industry without providing formal analysis. His main focus is on price discrimination among demanders (i.e. individual consumers) rather than among suppliers (i.e. hospitals, doctors in private practice and many other medical service providers). Kessel (1958) is a classic paper that deals with the price discrimination for medical care. It is, however, about price discrimination between wealthier versus poorer patients. Kessel's main argument is that doctors price discriminate simply because it is profitable rather than because they want to generously subsidise poorer patients. He also argues that even the free surgery they provide is to gain experience that enhances their skills. In our paper, we focus on price discrimination based on doctors' medical skills rather than price discrimination based on the patients' ability to pay.

Song *et al*. (2012) study inefficient price discrimination by private sellers of customised pharmaceutical 'cocktails' from which they derive market power, which contrasts with our focus on a price-discriminating public healthcare system. Chao's (2013) analysis of three-part tariffs may also be applicable to the way in which some healthcare systems (both private and public) price their services. Ye and Zhang (2017) find that optimal nonlinear pricing may depend crucially on whether consumers pay an entry cost before discovering their preference type, which could also have potentially important implications in the context of healthcare.

The paper is organised as follows. Section II provides the basic setup of the model. In section III, we develop the optimal medical fee schedule, which is characterised by differentiated medical fees based on medical cost. In section III, we consider policy implications for a national or other public healthcare system. Concluding remarks follow in section IV. All proofs are contained in the Appendix.

II. Basic Model

The government considers regulating the medical market which consists of a continuum of symmetric public healthcare providers. By 'symmetric', we mean that all healthcare service providers are equal in size. The cost of a typical hospital visit is represented by $C(q, \theta)$, where q is the quality of the medical service provided and θ is a parameter measuring (inversely) the type or ability of the medical service provider (e.g. a hospital or an individual doctor). For simplicity, we assume the service provider's technology can be described by the cost function $C(q, \theta) = \theta c(q)$, where $c' > 0$ and $c'' > 0$.

Based on the ability of the doctors and other staff employed by medical service providers, each provider in a population of providers is assumed to be either of low-cost (i.e. good) or high-cost (i.e. bad) type. Good-type hospitals, for example, hire high-ability doctors with low values of θ

(measuring inverse ability) and bad-type hospitals hire low-ability doctors with larger values of θ. In other words, good medical service providers are low-cost type, $\theta = \theta_L$, and bad medical service providers are high-cost type, $\theta = \theta_H$, where $\theta_L < \theta_H$.

A medical service provider's true type is its private information. The proportion of bad-type hospitals and doctors in the population of medical service providers is denoted $\lambda \in (0, 1)$, which is common knowledge. It is assumed that the level of quality, q, is observable and verifiable *ex post*, which implies that medical service is a so-called search good (i.e. its quality is easily evaluated prior to purchase) or an experience good (i.e. quality may be difficult to observe prior to purchase but can be ascertained with certainty after consuming the service) rather than a credence good (Nelson, 1970).[5]

We assume that there is no competition among medical service providers in the sense that medical fees are regulated by the government. We also assume that consumers are mandated to subscribe to the public healthcare service. Denoting the medical fee (i.e. the price of one standardised unit of medical service) as T, a medical service provider's profit function is $\pi = T - \theta c(q)$, and total consumer surplus is $CS = V(q) - T$, where $V(q)$ is total consumer utility when medical quality is q, such that $V' > 0$ and $V'' < 0$.[6] Here, we can interpret $V(q)$ as the sum of utilities across a population of heterogeneous consumers when medical service of quality q is provided.[7] The government is assumed to maximise social welfare (denoted W), which is defined as the sum of consumer surplus and a fixed share (denoted $\alpha \leq 1$) of hospitals' profits: $W = CS + \alpha\pi$. The parameter α reflects the extent to which the government cares about the profits of hospitals, doctors and other medical service providers.

Suppose that the government introduces a public healthcare system with medical fees differentiated by quality of the medical service provider. By the revelation principle, it suffices to restrict our attention to direct mechanisms. The government proposes a mechanism $\{q(\theta), T(\theta)\}_{\theta=\theta_L,\theta_H}$ to hospitals and doctors on a take-it-or-leave-it basis, where $q(\theta)$ and $T(\theta)$ represent medical quality and the fee (i.e. government transfer payment to hospitals and doctors) based on the message θ that the medical service provider sends to the government.[8] Let $q_H \equiv q(\theta_H)$, $q_L \equiv q(\theta_L)$, $T_H \equiv T(\theta_H)$ and $T_L \equiv T(\theta_L)$. The optimal mechanism solves the following constrained optimisation problem:

$$\max_{\{q(\theta),T(\theta)\}} E[W] \equiv \lambda[V(q_H) - T_H + \alpha\pi(\theta_H)] + (1-\lambda)[V(q_L) - T_L + \alpha\pi(\theta_L)], \tag{1}$$

[5] We exclude the case in which the quality of medical service remains unknown after patients receive service. If the quality of medical services cannot be learned or discovered even after consuming them, then medical services would be a so-called credence good. If medical service were a credence good, then fraud may occur. Although this possibility raises interesting modeling issues that would be worthwhile to explore in future research, the case of medical service as a credence good is beyond the scope of our analysis in this paper. Pitchik and Schotter (1987) and Wolinsky (1993) provide analyses of fraud in equilibrium.

[6] If consumers were not mandated to subscribe to the public healthcare system (contrary to what we have assumed), then the inequality $V(q) - T \geq 0$ would also need to be imposed as an additional constraint in the social-welfare maximization problem that follows, to guarantee that consumers prefer having the public healthcare service over not having it. Our main results (reported subsequently) showing efficiency gains when moving from uniform to quality-differentiated medical fees remain valid without imposing this constraint, however, as reductions in social inefficiency (under an unpopular public healthcare system that generates negative consumer surplus). This case should perhaps not be ruled out, given dissatisfaction about some healthcare systems expressed by many citizens.

[7] Formally, $V(q) = \int_{\tau \in \Upsilon} V(q; \tau)$, where τ is a parameter representing consumer preferences and $\tau \in \Upsilon$.

[8] One recent example of private medical service providers submitting messages (as represented by θ) to the public healthcare system would be the decision to join an ACO. Another example might be promises made by affiliated networks of private hospitals and insurers in the United States to participate in state healthcare exchanges under the ACA.

subject to:

$$T_H - \theta_H c(q_H) \geq 0, \qquad [IR_H]$$
$$T_L - \theta_L c(q_L) \geq 0, \qquad [IR_L]$$
$$T_H - \theta_H c(q_H) \geq T_L - \theta_H c(q_L), \qquad [IC_H]$$
$$T_L - \theta_L c(q_L) \geq T_H - \theta_L c(q_H). \qquad [IC_L]$$

The first two individual rationality conditions above ensure that both types of medical service providers will choose to participate in the quality-differentiated schedule of medical fees set by the price-discriminating public healthcare system. The other two constraints are incentive-compatibility conditions. [IC_H] requires that the high-cost-type service provider does not pretend being a low-cost type. Likewise, [IC_L] ensures that the low-cost-type service provider does not pretend being a high-cost type.

As usual, for a mechanism to be optimal, the bad type's participation constraint [IR_H] and the good type's incentive-compatibility constraint [IC_L] must be binding:

$$T_H = \theta_H c(q_H), \tag{2}$$

$$T_L = T_H + \theta_L c(q_L) - \theta_L c(q_H) = \theta_H c(q_H) + \theta_L c(q_L) - \theta_L c(q_H). \tag{3}$$

Note that $\partial C^2/\partial q \partial \theta = c' > 0$, implying that the single-crossing property holds. By virtue of this single-crossing property together with [IR_H] and [IC_L], then [IR_L] and [IC_H] are always satisfied. Thus, the optimisation problem given in (1) reduces to:

$$\begin{aligned} \max_{q_H, q_L} E[W] &= \lambda[V(q_H) - T_H + \alpha(\theta_H)] + (1-\lambda)[V(q_L) - T_L + \alpha\pi(\theta_L)] \\ &= \lambda[V(q_H) - \theta_H c(q_H) + \alpha\pi(\theta_H)] \\ &\quad + (1-\lambda)[V(q_L) - \theta_H c(q_H) - \theta_L c(q_L) + \theta_L c(q_H) + \alpha\pi(\theta_L)]. \end{aligned} \tag{4}$$

Using $\pi(\theta_H) = 0$ from [IR_H] and $\pi(\theta_L) = \theta_H c(q_H) - \theta_L c(q_H)$ from [IC_L], we have:

$$\frac{\partial E[W]}{\partial q_H} = \lambda[V'(q_H) - \theta_H c'(q_H)] + (1-\lambda)(1-\alpha)(\theta_H - \theta_L)c'(q_H) = 0, \tag{5}$$

$$\frac{\partial E[W]}{\partial q_L} = (1-\lambda)[V'(q_L) - \theta_L c'(q_L)] = 0. \tag{6}$$

The following proposition is the main result.

Proposition 1. *The optimal pricing mechanism under incomplete information achieves the following. (i) The good (low-cost) type of hospital with high-ability doctors provides the first-best socially optimal quality of medical service that would be chosen by a social welfare maximiser under complete information (denoted q_L^s for low-cost service providers): $q_L^* = q_L^s$. (ii) The bad (high-cost) type of hospital with low-ability doctors deliver worse quality than they would in a first-best social optimum under complete information: $q_H^* < q_H^s$. (iii) Low-cost types provide a higher quality of medical service than high-cost types do: $q_L^* > q_H^*$. And (iv) in return for providing* *higher quality, low-cost hospitals and doctors receive a higher medical fee than high-cost types:* $T_L^* > T_H^*$.

It is worthwhile to observe that the higher medical fee that low-cost hospitals and doctors receive includes an informational rent for good types as part of the social-welfare-maximising mechanism. They enjoy this informational rent due to the primary informational asymmetry motivating our analysis (i.e. hospitals and doctors know their own type but customers do not). In contrast, lower-quality hospitals receive only their reservation payoff. Is such a differentiated medical fees system necessary for achieving the social optimum?

We consider the cases of two extreme population distributions under uniform medical fee systems (as is actually the case in many countries' national healthcare systems). First, suppose that λ is so small that most of the population consists of low-cost hospitals. Then the government may prefer giving up a small proportion of the population of hospitals that are high-cost and instead pay a uniform (undistorted) medical fee only to low-cost hospitals.[9] In so doing, the government would disallow a positive surplus for the large proportion of hospitals that are good (i.e. low-cost types), refusing to pay a high-quality fee premium even though all hospitals choosing to operate in the market are good (or even excellent).[10] In this case, high-cost hospitals would choose to drop out of the market while only low-cost (good-type) hospitals would choose to deal with the government's optimal mechanism using first-best medical fees. In such a case, the government's maximisation of social welfare would lead it to essentially engage in a large-scale cream-skimming operation, effectively filtering out the small number of bad hospitals with low-skill staff, making them unable to profitably supply medical service given the government's low uniform fee structure.

Next we consider uniform pricing in the opposite case in which the population of hospitals is mostly bad and λ is therefore relatively large (i.e. most hospitals are high-cost types employing low-ability doctors and staff). Interestingly, in this case, the government does not need to provide a uniform high medical fee paid only to low-cost hospitals because the government already exploits them by allowing low-cost hospitals zero surplus as part of its optimal mechanism. Therefore, the government need not sacrifice good types in a uniform pricing rule when the population of hospitals is mostly bad.

[9] Would a government really consider closing hospitals that are sole providers of service to a particular geographic region? In fact, this kind of tension shows up in real-world policy debate and conflicts between local governments wanting to retain specialized services versus regional or federal governments wanting to close them. See, for example, debates over centralizing the provision of neurosurgery in Christchurch, New Zealand, which is five hours' drive from where those services are currently provided in Dunedin to residents in the Otago region. Accessed at https://www.odt.co.nz/news/dunedin/health/christchurch-neurosurgeon-denounces-propaganda

[10] New Zealand's healthcare system provides an interesting real-world case that, at least to a first approximation, plausibly fits with the theoretical case of uniform pricing in a population of hospitals that are mostly very good. New Zealand's doctors deliver excellent quality medical service that (despite problems) enjoys internationally benchmarked patient outcomes that rank among the best within the G30. New Zealand's public healthcare system constitutes a large share of total healthcare spending (well more than 80%, compared with just under 50% in the United States, based on OECD Health Data from 2011 as cited by the NZ Treasury, 2013). In international comparisons, New Zealand's healthcare system is championed for its focus on efficiency by aggressively managing costs, which includes notable success at negotiating substantially lower drug prices from international pharmaceutical suppliers than public healthcare systems in Europe or the US pay. Writing in the *Bulletin of the World Health Organization*, Lancashire (2010) describes how part of the rationalization for cost reductions and the public healthcare system's medical fee policies in New Zealand was to weed out hospitals employing low-skill staff.

We can draw some policy implications. In the first-best world in which the ability of doctors and hospital staff is heterogeneous and perfectly known to all consumers, it is obvious that medical fee discrimination is socially optimal. The analysis above shows that the same result of socially efficient price discrimination in government's setting of medical fees can be extended to the more realistic case in which hospitals, doctors and medical staff have private information about their ability unless the proportion of bad high-cost types is very low. It is interesting that in this case, good-type hospitals and doctors provide even higher quality than in the complete information case, because a good doctor wants to prevent bad-type hospitals and doctors from imitating them under incomplete information. Therefore, the institution of paying different medical fees for the same procedure to different hospitals and doctors based on their skill is socially desirable. Also, high fees can be interpreted as a signal for high quality.[11] Moreover, a good-type doctor enjoys a rent due to his or her informational advantage. Although the proportions of good-type and bad-type doctors are assumed to be fixed, the high rent that a good-type doctor earns provides a strong incentive for doctors to try hard to be good-type doctors. This is another social advantage of allowing differential medical fees.

III. A Continuum of Medical Service Provider Types

To make the model more realistic, in this section we assume that θ is distributed across the population of hospitals according to the density function $f(\theta)$ and the corresponding distribution function $F(\theta)$ on an interval $\Theta \equiv \left[\underline{\theta}, \overline{\theta}\right]$, where $\underline{\theta} < \overline{\theta}$. We assume that the inverse hazard rate $\frac{F(\theta)}{f(\theta)}$ is increasing in θ, as usual in the mechanism design literature.

Again, we can use the revelation principle to restrict attention to direct mechanisms: $\{q(\theta), T(q(\theta))\}_{\theta\in\Theta}$. Suppose a doctor with θ provides the quality $q(\theta)$ and gets paid $T(q(\theta))$. Social welfare can be written as:

$$W = \int_{\underline{\theta}}^{\overline{\theta}} (V(q(\theta)) - T(\theta) + \alpha\pi(\theta)) f(\theta) d\theta. \tag{7}$$

The individual rationality conditions and incentive-compatibility conditions are:

$$T(q(\theta)) - \theta c(q(\theta)) \geq 0, \forall\theta \in \Theta, \qquad [PC]$$

$$T(q(\theta)) - \theta c(q(\theta)) \geq T\left(q\left(\widetilde{\theta}\right)\right) - \theta c\left(q\left(\widetilde{\theta}\right)\right), \forall\theta, \forall\widetilde{\theta} \neq \theta, \qquad [IC]$$

where $\widetilde{\theta}$ is interpreted as each respective hospital's (or doctor or other medical service provider's) self-reported type. We use the first-order approach. Assuming that $q' > 0$ and $T' > 0$, we can replace the[*IC*] condition by the following *local* incentive-compatibility condition [*LIC*]:

$$T'(q(\theta)) = \theta c'(q(\theta)). \qquad [LIC]$$

The incentive-compatibility condition [*IC*] also implies that:

[11] This does not quite fit the conventional meaning of a signal, which is an (observable) action of an informed player who chooses his action before an uninformed player chooses his own action.

$$\max_{\tilde{\theta}} T\left(q\left(\tilde{\theta}\right)\right) - \theta c\left(q\left(\tilde{\theta}\right)\right) = T(q(\theta)) - \theta c(q(\theta)) \equiv \pi(\theta). \tag{8}$$

Because equation (8) must hold for all $\theta \in \Theta$, we have (by the Envelope Theorem):

$$\pi'(\theta) = -c(q(\theta)). \tag{9}$$

Integrating equation (9), we obtain:

$$\pi(\theta) = \pi\left(\overline{\theta}\right) - \int_{\underline{\theta}}^{\overline{\theta}} \pi'(t)dt = \pi\left(\overline{\theta}\right) + \int_{\underline{\theta}}^{\overline{\theta}} c(q(t))dt = \int_{\underline{\theta}}^{\overline{\theta}} c(q(t))dt. \tag{10}$$

Using $\pi\left(\overline{\theta}\right) = 0$, we then have the following:

$$\pi(\theta) = T(q(\theta)) - \theta c(q(\theta)) = \int_{\underline{\theta}}^{\overline{\theta}} c(q(t))dt. \tag{11}$$

Substituting equation (11) into the objective function yields:

$$W = \int_{\underline{\theta}}^{\overline{\theta}} \left(V(q(\theta)) - \theta c(q(\theta)) + (\alpha - 1) \int_{\theta}^{\overline{\theta}} c(q(t))dt \right) f(\theta)d\theta. \tag{12}$$

Integrating (12) by parts provides the following:

$$\int_{\theta}^{\overline{\theta}} \left(\int_{\theta}^{\overline{\theta}} c(q(t))dt \right) f(\theta)d\theta = \left[\int_{\theta}^{\overline{\theta}} c(q(t))dtF(\theta) \right]_{\underline{\theta}}^{\overline{\theta}} + \int_{\theta}^{\overline{\theta}} c(q(\theta))F(\theta)d\theta = \int_{\theta}^{\overline{\theta}} c(q(\theta))F(\theta)d\theta, \tag{13}$$

because $\left[\int_{\theta}^{\overline{\theta}} c(q(t))dtF(\theta) \right]_{\underline{\theta}}^{\overline{\theta}} = 0$. Therefore, the socially efficient choice of $q(\theta)$, denoted $q^*(\theta)$, is the solution to the problem of maximising the following social welfare objective:

$$W(q(\theta)) = \int_{\underline{\theta}}^{\overline{\theta}} [(V(q(\theta)) - \theta c(q(\theta)))f(\theta) + (\alpha - 1)c(q(\theta))F(\theta)]d\theta. \tag{14}$$

The integrand above is referred to as *virtual social welfare.* The first term is the first-best social welfare and the second term is the negative effect of the incentive problem on social welfare. It is to keep the informational rent decreasing in θ. Differentiating the expression above with respect to $q(\theta)$, we get:

$$[V'(q(\theta)) - \theta c'(q(\theta))]f(\theta) + (\alpha - 1)c'(q(\theta))F(\theta) = 0, \tag{15}$$

or, equivalently:

$$V'(q(\theta)) = \theta c'(q(\theta)) + (1 - \alpha)c'(q(\theta))\frac{F(\theta)}{f(\theta)}. \tag{16}$$

The equation above requires that the consumer's marginal benefit must be equal to the medical service provider's marginal *virtual cost.* The virtual cost of a given type takes into account not only own production cost but also the cost of deterring other (more efficient types) types from imitating the service provider's own type. Because the cost of deterring better types from imitating which is measured by the second term on the right-hand-side is positive, it implies that informational rent is in fact decreasing in θ, such that a doctor with a lower value of θ cannot imitate one with a higher value of θ. As a result, the allocation is no longer efficient as it was under complete information.

Equation (16) states that, under incomplete information, there is a fundamental tradeoff between implementing allocations close to efficiency and compensating the most efficient types with informational rents to induce information revelation. This tradeoff prescribes distortions away from first-best social efficiency with complete information.

From the [*LIC*] condition, it also follows that:

$$T'(q(\theta)) = \theta c'(q(\theta)). \tag{17}$$

The following proposition generalises Proposition 1 to the case of a continuum of medical service provider types by characterising the θ-type level of quality that the socially efficient mechanism under incomplete information achieves ($q^*(\theta)$) relative to the first-best socially efficient level of quality under the counterfactual of complete information (denoted $q^S(\theta)$).

Proposition 2. *By allowing for medical fee price discrimination, the socially efficient mechanism under incomplete information is characterised by the following: (i)* $q^*(\theta) < q^S(\theta)$ *for all* $\theta \neq \underline{\theta}$*; (ii)* $q^*(\theta)$ *is decreasing in* θ*; and (iii)* $T(q)/q$ *is increasing in q.*

Proposition 2 says that, under incomplete information, all hospitals and doctors—except for those with the very highest ability—provide lower quality than would be prescribed in a first-best social optimum under complete information. The highest-ability medical service provider achieves the first-best socially efficient quality and enjoys the largest informational rent paid out by the national healthcare system. This medical fee should be large enough to prevent him from imitating low-ability types by providing a lower quality service while earning lower medical fees per service. Informational rents paid to medical service providers decrease for those of lower ability (higher θ) and converge to zero rents for the hospital or doctor with the lowest ability.

IV. Conclusion

In this paper, we showed that introducing medical fee differentiation by different tiers of quality among medical service providers can substantially improve social efficiency. We argued that doing so has an additionally beneficial dynamic effect by incentivising hospitals, doctors and other medical services providers to improve their ability levels and thereby enjoy higher informational rents. We believe that our results can be easily extended to the case of private health insurers to hospitals and doctors according to their quality of service could achieve similar efficiency improvements, unless they insist on 'managed care' whereby choice of physicians, health care facilities and the amount charged for the services are limited, as in the United States under the ACA.

APPENDIX

Proof of Proposition 1.

(i) In the first-best world, (IR_H) and (IR_L) are binding, that is, $T_H = \theta_H c(q_H)$ and $T_L = \theta_L c(q_L)$. Thus, the social optimum is

$$\max_{\{q(\theta)\}} E[W] \equiv \lambda[V(q_H) - \theta_H c(q_H) + \alpha\pi(\theta_H)] + (1-\lambda)[V(q_L) - \theta_L c(q_L) + \alpha\pi(\theta_L)].$$

The first-order condition requires $V'\left(q_H^S\right) = \theta_H c'\left(q_H^S\right)$ and $V'\left(q_L^S\right) = \theta_L c'\left(q_L^S\right)$. Also, from equation (6), we have $V'\left(q_L^*\right) = \theta_L c'\left(q_L^*\right)$, implying that $q_L^* = q_L^S$.

(ii) It follows from equation (5) that $V'\left(q_H^*\right) = \theta_H c'\left(q_H^*\right) + \Delta$ where $\Delta \equiv \frac{1-\lambda}{\lambda}(1-\alpha)(\theta_H - \theta_L)c'\left(q_H^*\right)$. Because $\Delta > 0$, we have $q_H^* < q_H^s$ from $V''(\cdot) < 0$ and $c''(\cdot) > 0$.

(iii) Differentiating the first-order conditions yields $V''dq = \theta c''dq + c'd\theta$, that is, $\frac{dq}{d\theta} = \frac{c'}{V'' - \theta c''} < 0$. Therefore, it follows that $q_H^* < q_H^S < q_L^S = q_L^*$.

(iv) We have $T_L^* - T_H^* = \theta_L\left(c\left(q_L^*\right) - c\left(q_H^*\right)\right) > 0$, because $q_L^* > q_H^*$.

Proof of Proposition 2.

(i) Any social optimum requires that:

$$V'\left(q^S(\theta)\right) = \theta c'\left(q^S(\theta)\right). \tag{A1}$$

Comparing equation (A1) with equation (15) implies that $q^*(\theta) < q^S(\theta)$ for all $\theta \neq \underline{\theta}$. If $\theta = \underline{\theta}$, $F\left(\overline{\theta}\right) = 0$ so that $q^S\left(\underline{\theta}\right) = q^*\left(\underline{\theta}\right)$.

(ii) Total differentiation of equation (15) yields

$$[V'' - \theta c'' + (\alpha - 1)c''\psi]dq + [-c' + (\alpha - 1)c'\psi']d\theta = 0,$$

where $\psi(\theta) = \frac{F(\theta)}{f(\theta)}$. Because $V'' < 0$, $c'' > 0$ and $\alpha \leq 1$, we have $\frac{dq}{d\theta} < 0$.

(iii) Differentiating equation (17) leads to $T'' = \theta c'' > 0$. Because $T(q)$ is convex, $T(q)/q$ is increasing in q.

REFERENCES

Allin, S., Davaki K. and Mossialos, E. 2006, ‘Paying for ‘Free’ Health Care: The Conundrum of Informal Payments in Post-Communist Europe’, *Transparency International Global Corruption Report*, pp. 62–75.

Atanasova, E., Pavlova, M. and Groot, W. 2015, ‘Out-of-Pocket Patient Payments for Public Health Care Services in Bulgaria’, *Frontiers in Public Health*, vol. 3, p. 175.

Balabanova, D. and McKee, M. 2002, ‘Understanding Informal Payments for Health Care: The Example of Bulgaria’, *Health Policy*, vol. 62, pp. 243–273.

Chao, Y. 2013, ‘Strategic Effects of Three-Part Tariffs Under Oligopoly’, *International Economic Review*, vol. 54, pp. 977–1015.

Chawla, M., Berman, P. and Kawiorska, D. 1998, 'Financing Health Services in Poland: New Evidence on Private Expenditures', *Health Economics*, vol. 7, pp. 337–346.

Courtemanche, C., Marton, J. and Yelowitz, A. 2014, 'Who Gained Insurance Coverage in 2014, the First Year of Full ACA Implementation?', *Health Economics*, vol. 25, pp. 1099–1050.

Ensor, T. 2004, 'Informal Payments for Health Care in Transition Economies', *Health Policy*, vol. 58, pp. 237–246.

Goldman, M.B., Leland, H. and Sibley, D. 1984, 'Optimal Nonuniform Prices', *Review of Economic Studies*, vol. 51, pp. 305–319.

Kessel, R. 1958, 'Price Discrimination in Medicine', *Journal of Law and Economics*, vol. 1, pp. 20–53.

Lancashire, R. 2010, 'New Zealand Cuts Health Spending to Control Costs', *Bulletin of the World Health Organization*, vol. 88, pp. 717–796.

Lewis, M.A. 2000, *Who Is Paying for Health Care in Eastern Europe and Central Asia?* Washington, DC, World Bank Publications.

Maskin, E. and Riley, J. 1984, 'Monopoly with Incomplete Information', *RAND Journal of Economics*, vol. 15, pp. 171–196.

Matthews, S. and Moore, J. 1987, 'Monopoly Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening', *Econometrica*, vol. 55, pp. 441–467.

Miller, W.L. and Koshechkina, T.Y. 2000, 'If You Pay, We'll Operate Immediately', *Journal of Medical Ethics*, vol. 26, pp. 305–311.

Mirman, L. and Sibley, D. 1980, 'Optimal Nonlinear Prices for Multiproduct Monopolies', *Bell Journal of Economics*, vol. 11, pp. 659–670.

Mussa, M. and Rosen, S. 1978, 'Monopoly and Product Quality', *Journal of Economic Theory*, vol. 18, pp. 301–317.

Nelson, P. 1970, 'Information and Consumer Behavior', *Journal of Political Economy*, vol. 78, no. 2, pp. 311–329.

OECD 2005, *Economic Surveys: Hungary 2005*, Paris, Organisation for Economic Co-operation and Development Publishing.

Phlips, L. 1983, *The Economics of Price Discrimination*, Cambridge, MA, Cambridge University Press.

Pitchik, C. and Schotter, A. 1987, 'Honesty in a Model of Strategic Information Transmission', *American Economic Review*, vol. 77, pp. 1032–1036.

Reinhardt, U. 2006, 'The Pricing of U.S. Hospital Services: Chaos behind a Veil of Secrecy', *Health Affairs*, vol. 25, pp. 57–67.

Sharkey, W. and Sibley, D. 1993, 'Optimal Non-Linear Pricing with Regulatory Preference over Customer Type', *Journal of Public Economics*, vol. 50, pp. 197–229.

Song, M., Nicholson, S. and Lucarelli, C. 2012, *Bundling Among Rivals: A Case of Pharmaceutical Cocktails*', Working Paper, Rochester, NY, University of Rochester.

Srinagesh, P. and Bradburd, R. 1989, 'Quality Distortion by a Discriminating Monopolist', *American Economic Review*, vol. 79, pp. 96–105.

Stole, L. 2007, 'Price Discrimination and Competition', in M. Armstrong and R. Porter (eds), *Handbook of Industrial Organization*, vol. 3, New York, North-Holland, pp. 2221–2299.

NZ Treasury 2013, *Health Projections and Policy Options for the 2013 Long-Term Fiscal Statement*, Wellington, Treasury Publications, http://www.treasury.govt.nz/government/longterm/externalpanel/pdfs/ltfep-s4-01.pdf

Wolinsky, A. 1993, 'Competition in a Market for Informed Experts' Services', *RAND Journal of Economics*, vol. 24, pp. 380–398.

Ye, L. and Zhang, C. 2017, 'Monopolistic Nonlinear Pricing with Consumer Entry', *Theoretical Economics*, vol. 12, no. 1, pp. 141–173.