

Stanford Center for International Development

Working Paper No. 373

Incentives in China's Healthcare Delivery System

by

Karen Eggleston

October 2008

Stanford University

579 Serra Mall @ Galvez, Landau Economics Building, Room 153

Stanford, CA 94305-6015

INCENTIVES IN CHINA'S HEALTHCARE DELIVERY SYSTEM

Karen Eggleston[†]

October 2008

Abstract

The economic approach of comparative and historical institutional analysis (Aoki 2001, Greif 2006) has virtually never been used in theoretical studies of healthcare incentives. This paper seeks to help fill this gap by exploring the explanatory power of such an approach for understanding incentives in China's healthcare delivery system. It focuses on positive analysis of why China's health system incentives evolved the way they did. The first section analyzes the institution of physician dispensing (MDD) and reforms toward separation of prescribing from dispensing (SPD), in historical and comparative perspective. It shows, for example, how MDD was a self-reinforcing institution; the longer a society remains under MDD, the higher the associated costs of supplier-induced demand can be before implementing SPD becomes the efficient self-enforcing social institution. Rapid technological change and adoption of universal coverage are likely to trigger SPD reforms. The second section seeks to explain the pattern and impact of price regulation and hospital payment reforms in contemporary China, which also reflect the legacy of MDD.

Keywords: Healthcare, China, public policy.

JEL Classification No.: H51, I11, I18.

[†] Stanford University Shorenstein Asia-Pacific Research Center, Encina Hall E301, Stanford, CA 94305-6055, corresponding email: karene@stanford.edu.

1 Introduction

In health care, as in other arenas of human endeavor, individuals and organizations respond to the institutional structure of incentives. Society prospers when organizations have incentive to invest in productive outputs and innovations (North 2005). To understand the adaptive efficiency of health care systems, it can be valuable to consider economic incentives in their historical context, drawing from related contributions in anthropology, sociology, medicine, public health, and area studies. Yet this economic approach of comparative and historical institutional analysis (Aoki 2001, Greif 2006) has virtually never been used in theoretical studies of healthcare incentives.

This paper seeks to help fill this gap by exploring the explanatory power of such an approach for understanding incentives in China's healthcare delivery system. How are China's current payment systems and their reforms conditioned by the institutional elements – beliefs, norms, and expectations – inherited from the past (Greif 2006), such as traditional Chinese medicine (TCM), physician dispensing, and economic transition? Why do hospitals in China earn more than 40 percent of their revenues from drug dispensing to patients? Why did policymakers set regulated prices such that drug sales are profitable, but diagnostic and management services were not? Why has separation of prescribing from dispensing become central to the policy debate recently? Why have some hospitals in China voluntarily adopted case-based payment? In short, what factors explain the pattern of China's healthcare incentives and the process of China's payment reforms?

This paper seeks not to judge or advise, but rather to explain; the intent is positive rather than normative analysis. Since it would be impossible to cover the whole array of healthcare delivery incentives in one model or even series of models, I focus instead on two key examples: (1) The institution of physician dispensing and reforms toward separation of prescribing from dispensing (SPD) in China, in historical and comparative perspective; and (2) hospital payment reforms, especially the current widespread experimentation with case-based payment, global budgets, and 'separation of revenues and expenditures.' Together these analyses illustrate how the integration of prescribing and dispensing became an embedded feature of China's current delivery system, and how such arrangements continue to have a significant impact on policy reforms.

2 Physician-Dispensing as a Social Institution

Greif (2006, p.30) defines an institution as a system of rules, beliefs, norms and organizations that together generate a regularity of social behavior. This section explores the theory of physician-dispensing as a social institution and the comparative historical evolution of SPD policies. In part this is an inquiry into what Abbott (1988) calls “the ecology of professions” in comparative sociology, exploring the boundaries between doctors and pharmacists. In part it is also an application to health care of the economic theory of expert-client interactions when information is asymmetric (Wolinsky 1993; Taylor 1995; Hubbard 2002; Afendulis and Kessler 2007). Most centrally, however, it is an inquiry applying to health systems the method of comparative and historical institutional analysis (Aoki 2001, Greif 2006). We seek to explain physician dispensing as a social institution, or a regularity of doctor-patient behavior.

Under what conditions were diagnosis and evaluation (Dx), on the one hand, and prescription and dispensing of medication (Rx), on the other, integrated within a single provider rather than separated between a physician and an independent apothecary or pharmacy? With this way of framing the question, it is natural to seek conceptual answers in the literature on relational contracting and the boundaries of the firm. In particular, I modify and apply to the SPD question the theory of asset ownership and relational contracting developed by Baker, Gibbons and Murphy (2002; hereafter BGM).

Relational contracts are “informal agreements sustained by the value of future relationships” (BGM, p. 39). Since the central transaction with which I am interested is that between physician and patient, I consider relational contracts between a patient/purchaser and a provider. Relational contracts provide incentives for noncontractible commitments (provider efforts, purchaser future business and referrals) if the parties sufficiently value their reputations and the future benefits associated with the doctor-patient relationship.

A physician diagnosing a patient who also derives income from selling treatment (such as a procedure or medication) may take advantage of asymmetric information to recommend treatments that would not be recommended in the absence of such an incentive. A doctor-dispenser has incentive and opportunity for “supplier induced demand” (SID; see for example Gruber and Owings 1996) and/or monopoly pricing of an innovative drug or therapy. Therefore the current incentives and constraints on SID and pricing – such as whether the doctor or hospital must give the patient

a written prescription and cannot hire a pharmacist or own a pharmacy – play an important role in determining whether the parties will honor the relational contract.

I first provide some historical background, and then proceed to develop a model of physician dispensing that can help to explain the path dependence of China’s incentive structure, while rejecting the simplistic view that integration is backward and separation modern.

2.1 Historical background

The cultures of healing and professions of medicine in the east and the west differed in many important ways, and each is itself a heterogeneous collection of many schools of thought. Here I focus on the role of medications in the patient-physician interaction. In the west, prescribing and dispensing of medications has often been ancillary to the primary expected outcomes of a doctor visit, diagnosis and explanation.¹

In herbal medical traditions as dominated in East Asia, however, the prescription and preparation of medications was central to the entire enterprise, and the dispensing of medicine came to be seen as the central outcome of the patient-physician interaction. For example, from ethnographic study as an anthropologist and clinician, Kleinman (1980, pp.260-264) describes clinical interactions between patients and Chinese-style doctors that feature little or no explanations, and belief that the doctor’s therapeutic power stems from emotional distance from the patient and the detailed prescription. He notes that secrecy is a central element of Chinese medicine, with emphasis on herbal medicines and diet as the essence of the Chinese medical tradition (ibid, pp.275-277).

Unschuld (1986) has estimated that “works meant to impart knowledge about individual drugs” (p.2) alone constitute about 14% of all medical-related literature in China from over more than a thousand years; and works specific to prescription are another large category. This in itself points to the important role of pharmaceuticals in TCM. Few other societies saw the establishment of a

¹As Roy Porter quotes in his opus *The Greatest Benefit to Mankind: A Medical History of Humanity from Antiquity to the Present*, distinguished American physician Lewis Thomas, who was a medical student in the 1930s, recalls

we were provided with a thin, pocket-sized book called *Useful Drugs*, one hundred pages or so, and we carried this around in our white coats when we entered the teaching wards and clinics in the third year, but I cannot recall any of our instructors ever referring to this volume. . . . *Our task for the future was to be diagnoses and explanation. Explanation was the real business of medicine.* What the ill patient and his family wanted most was to know the name of the illness, and then, if possible, what had caused it, and finally, most important of all, how it was likely to turn out (Porter 1997, pp.681-2, italics added).

national pharmaceutical association even before that of the professional association for physicians.² Under the influence of Japan's Meiji Restoration, a movement in China arose to get rid of or 'at least modernize' traditional medicine, setting up an adversarial relationship between western biomedicine and TCM (*Shijie Yiyao Weisheng 100 Nian*, p.14). However, the founders of the PRC rejected this approach as a 'false dichotomy.' The integration of western and Chinese medicine, a characteristic of the PRC system from its founding to today, developed partly in recognition of the reality of many TCM doctors and few western doctors relative to the population. Mao Zedong was an advocate of integrating TCM and western medicine, using the latter science to discover the traditional wisdom passed down through generations of TCM practice (*ibid*, pp.14-15). In 1982 development of TCM became part of the PRC constitution (*ibid*), and the current five-year plan emphasizes innovation and development of TCM and biopharmaceuticals.

In the tradition of TCM, also shared by Japan and Korea, 'doctor-dispensers' and 'druggist-prescribers' have long co-existed and competed for patients. Indeed, it was considered improper for doctors to charge patients for the humane service of diagnosis, so that doctors primarily made a living by selling medications (Ikegami in Gauld 2005). This pattern became a part of the modern healthcare system as well. As Eggleston and Yang (2008) note, "pharmacists in Korea essentially played the role of primary care giver for nearly four and a half decades. They responded to medical inquiries by making a diagnosis, and then dispensing drugs. Most Koreans accepted such a tradition, and it became a large facet of what health care was to them."

In the early 1950s at the outset of the PRC, the tradition of physician-dispensing found a modern counterpart in the decision to allow hospitals to charge a mark-up for selling pharmaceuticals (Zhang 2008). Although the policies regarding hospital finance and pharmaceutical price regulation have changed over the years, this mark-up for dispensing drugs has remained (Huang and Yang 2008). Pharmaceutical prices have long been regulated in China, except from 1992 to 1996, when the Chinese government let the market set drug prices (Sun et al. 2008). Pharmaceuticals account for about half of total healthcare expenditures in China, representing 43% of expenditure per inpatient episode and 51% of expenditure per outpatient visit (*ibid*). This relatively large share appears in part related to hospitals or other providers receiving between 15% (the official mark-up) to 40% or more of the retail price of pharmaceuticals. Indeed, if the provider margin is simply reallocated from

²The Chinese Pharmaceutical Association (Zhongguo Yaoxue Hui) was founded in 1907 by a Chinese scholar studying in Japan and moved its headquarters back to Beijing in 1912. The professional organization of Chinese physicians, the Chinese Medical Association (Zhonghua Yixue Hui) was established in 1915.

pharmaceutical spending to providers, the distribution of healthcare expenditures across providers, pharmaceuticals and other services is much more similar to those of other middle-income countries.

But the tradition has ramifications beyond simple accounting. “Supporting hospitals through drug sales” (*yi yao yang yi*) is widely recognized as a problem. The brief account of its genesis and contemporary manifestations given by Sun and colleagues (2008) is worth quoting at length:

Government financial support to public hospitals declined significantly in the reform era. At the beginning of the 1980s, state financing constituted about 60% of hospital revenues, but by 2003 it had fallen to 8.2%. To make up for the shortfall, hospitals have come to regard pharmaceutical sales as one of their main sources of revenue. From the mid-1950s to the present, the government has officially permitted a 15% mark-up for hospitals. In 2006, the NDRC reiterated that retail prices should equal the provider’s procurement price plus a 15% profit margin. However, some evidence suggests that the actual mark-up of medicines in government hospitals in 2005 was much higher, at about 42%. Even after years of effort to reduce hospital reliance on drug sales to finance their operating costs, for the average hospital in China in 2006 drug sales accounted for 41.5% of total income. At the same time drug purchase costs accounted for 37.8% of hospital expenses, which indicates the importance of drug sales for a hospital’s bottom line....Hospitals account for roughly four-fifths of all retail pharmaceutical sales. One study of seven public hospitals at and over the county level found that only 29% of outpatients filled their prescriptions in retail drug stores rather than at the treatment hospital’s pharmacy. (Sun et al. 2008)

Why did “*yi yao yang yi*” arise in China and many East Asian neighbors, but not elsewhere? The next sections attempt to build a simple general theory that can illuminate this question.

2.2 Institutional arrangements regarding prescribing and dispensing

Interactions between doctors and patients take place within a given historical and cultural context. “Social reality is constructed or created in the sense that certain meanings, social structural configurations, and behaviors are sanctioned (or legitimized), while others are not” (Kleinman 1980, p.36). Social reality as applied to the clinical encounter between medical provider and patient can be called “clinical reality” (ibid). Understanding under what circumstances clinical reality, based on patient

and provider beliefs and norms, embraces or eschews separation of diagnosis and dispensing should provide insight for understanding both western biomedicine and Asian herbal medicine traditions and the path dependence of current pharmaceutical policies and broader incentive structures.

Consider two primary institutional arrangements governing clinical reality in a given society at a given time: (1) integration of Dx and Rx, termed ‘MD-dispensing’ (MDD); and (2) strict enforcement of patients’ property rights to the prescription, so that the patient may have it filled at any pharmacy and the provider does not receive dispensing revenues. This latter arrangement, familiar to many in the west, is denoted ‘separation of prescribing and dispensing’ (SPD). I say that an arrangement prevails as a social institution when both parties – the upstream provider and the downstream purchaser – find the value of future interactions under the current institutional arrangement outweighs the temptation to renege on the relationship (and resort to the spot market or, eventually, change the social institution).

MDD can arise in a society in which no legal or contracting institutions enforce restrictions on the physician-patient interaction or on the property rights to a given medication. This can be thought of as the original situation in all societies, and has several implications for Dx and Rx behavior. First, patients can choose to buy drugs of any form in any place where they are marketed: druggists, pharmacies, or doctor offices. Second, if a doctor or druggist develops a new medical formulation that effectively treats a given medical condition, that innovator has no enforceable property rights to the formulation. Other druggists, doctors, and patients can expropriate the innovation, unless the originator keeps it secret by integrating the diagnosis (Dx) and dispensing (Rx) functions (by, for example, hiring relatives as dispensers). Third, there is no distinction between prescription and over-the-counter drugs. Fourth, doctors can freely choose whether to specialize in diagnosis or to integrate with a druggist and also dispense medications. As we will show, they would typically choose integration unless SPD is enforced. Indeed, under MDD there may be few licensing and accreditation standards for providers; itinerant doctors selling ‘patent medicines’ may charge what the market will bear, and compete with druggists and pharmacists who diagnose and prescribe as well as dispense medications. In such an environment, MDD allows the physician to claim ownership over innovative medication by keeping the ingredients secret (*mifang*). The linking of Rx with a physician also provides some reputational mechanism to guarantee quality of the medication.³

³Interestingly, an article written in 809 by a local Chinese government official in southern China recounts how he was cheated when given a prescription and buying ingredients: A doctor had apparently written the official

In contrast, some societies have adopted restrictive laws that enforce a division of labor between physicians and pharmacists, property rights over drug discovery, a distinction between prescription-only and over-the-counter medications, as well as legally restrict physician or hospital ownership of ancillary medical facilities such as pharmacies. Only when a society has invested in such laws and enforcement institutions can SPD prevail. In South Korea, for example, separation reforms in 2000 forced hospitals to shut down and divest from all outpatient pharmacies.

I denote all contracting costs associated with SPD as Ω . Although intermediate institutional arrangements may exist, such as restrictions on ownership that are not enforced or even flagrantly violated, I focus here on MDD and SPD to illustrate the two dominant extremes.

The key question is, when will parties find MDD or SPD the preferable self-enforcing structure governing the doctor-patient relationship? To move from MDD to SPD requires paying $\Omega > 0$. Patients also must separately visit both the doctor and a pharmacy, making two (or more) trips instead of one for each treatment episode. Let patients' perception of the differential travel cost associated with SPD, at time t , be T_t . (If travel costs under MDD are T^L and under SPD are T^H , then $T = T^H - T^L$).

Travel costs, although often not highlighted in discussions of medical care in industrialized countries, can be and have been in the past a very significant component of the costs of medical practice and of alternative organizational arrangements. Paul Starr (1980) notes, for example, how indirect fees for travel constituted 50 percent or more of medical fees in the nineteenth century United States.⁴

a prescription for an unspecified illness. After taking the requisite medicine bought at a local market, however, [he] felt even worse than before. His doctor subsequently examined the dregs of the decoction and found that one of the traditional medicines he had been prescribed...was not the genuine article but rather a piece of taro. (Du 2005, p.38).

This historical account provides evidence that historically there was a time when patients in China received prescriptions from doctors and purchased the medicine (ingredients) elsewhere; and second, illustrating the pros and cons of separation in traditional society—ability to shop around for ingredients and prepare own medicine, but risk of low quality, such as ingredients that are not authentic.

⁴For example, in Mississippi in 1843, a doctor's visit cost \$1 and the charge per mile of travel was \$1 (\$2 at night).

These ratios between charges for services and mileage are typical. Even at relatively short distances, the share of the total price due to traveling and opportunity costs exceeded the physician's ordinary fee; at a distance of five or ten miles, the mileage charges typically amounted to four or five times the basic fee for a visit [Starr 1980, p.67]....The telephone made it less costly to reach a physician by greatly reducing the time formerly tracking down the peripatetic practitioner on foot. Phones first became available in the late 1870s. Curiously, the first rudimentary telephone exchange on record, built in 1877, connected the Capital Avenue Drugstore in Hartford, Connecticut, with twenty-one local doctors [Starr 1980, p.69].

Travel costs and the associated opportunity cost of time for patients remain significant components of the costs of health care, especially in developing and middle-income countries like China.⁵ Moreover, consumers' perception of the costs associated with the need to visit doctors and pharmacies on separate trips may differ according to how long they have experienced integration (MDD) or separation (SPD). The perceived cost is likely to be largest, for example, if consumers have lived with integration for many generations. Therefore, I make the following assumption:

Claim 1 T_t increases with the length of time under MDD.

MDD avoids the contracting costs Ω and the travel or transaction costs T_t , but can be associated with inefficient incentives for doctors to over-prescribe and dispense medications according to profitability. I denote this potential welfare loss from MDD as 'supplier-induced demand' (SID), using notation P . As we will see, encouraging an unproductive action such as pure SID reduces total surplus but may also reduce the renegeing temptation in the doctor-patient relationship enough to make it worthwhile (for a similar argument see BGM 2002, p.67). Moreover, committing to reward doctor prescribing, even with prices reflecting market power, can serve as a relational incentive for quality effort, when doctors value future reputation. Indeed, when contractings costs Ω are high, rewarding provider innovation and assuring patients of quality are both central concerns which integration of diagnosis and treatment (MDD) may help to address. If the doctor and hospital make profits and reputation from sales of quality drugs, then MDD also rewards their efforts to monitor quality of drugs and reduce fake medicines. This already hints at one of the possible advantages of MDD in China, given the considerable concern there regarding the safety and quality of drugs and other items (epitomized by the recent baby formula and milk contamination; also see Santoro and Liu 2008).

BGM note that "the key question in our analysis is whether choosing appropriate asset ownership (integration or nonintegration) can make a given promise self-enforcing" (p.41). Here, the key question is whether physician or patient ownership of the Rx "asset" (rights to where to fill the prescription and the associated revenue) can make two promises self-enforcing: (a) the purchaser promise not to expropriate the *mifang* of a doctor-inventor or 'defect' from the doctor-patient re-

⁵Direct evidence is scanty, but it is quite plausible that such travel costs are particularly significant in China's rural areas and remoter part of western China. Moreover and perhaps less obviously, the opportunity cost of patients' time can also be a significant portion of total burden for urban patients, who must queue in sometimes long lines for high-reputation hospitals and doctors. As the popular lament goes (which is even mentioned in China's current 5-year plan), seeing a doctor is not only expensive, but difficult (*kan bing nan, kan bing gui*).

relationship to self-diagnose (perhaps with pharmacists ‘quasi-prescribing’ thereafter); and (b) the physician promise of not (excessively) inducing demand.

The “temptation to renege” is important because while MDD may be an institutional equilibrium in the overall society, the meeting of any one patient with any one doctor may be infrequent; therefore the doctor’s ability to induce demand (including the patient’s ‘gullibility’ and willingness and ability to pay given the severity of the condition) will vary across encounters. When many doctor-patient interactions lead to temptation to renege on relational contracts, then the institution governing the doctor-patient relationship (MDD or SPD) begins to break down and is no longer self-enforcing.

MDD or SPD may foster different intensities or forms of competition for providers (such as MD-dispensers with prescribing pharmacists under MDD, compared to separate markets for physicians and independent pharmacies under SPD). However, to focus on the impact of organizational arrangements of MDD and SPD, we abstract from competition to focus on the interaction between a representative patient and provider.

2.3 Elements of a historical comparative theory of physician dispensing

The upstream party, the doctor, produces Dx services which, combined with the filled prescription (Rx), treat the downstream patient’s medical condition. If the doctor owns the Rx asset, then MDD prevails; BGM call this ‘relational outsourcing.’ If the downstream patient owns the Rx asset, then SPD prevails; BGM call this ‘relational employment.’ I modify the model to fit MDD and SPD (e.g., including SID).

2.3.1 MD efforts

Each period, the doctor chooses two actions, quality-related effort e_1 and effort to induce demand, e_2 , at cost $c(e_1, e_2)$. No one except the doctor herself can observe the actions e_1 and e_2 .

Patient benefit from doctor quality effort, Q_i , can take two values: Q_o or Q_H . The minimum contractible quality is normalized to zero: $Q_o \equiv 0$. With probability $q(e_1)$, patient benefit increases to $Q_H > 0$. In other words, the surplus associated with doctor quality effort is probability $q(e_1)$ of benefit $Q_i = Q_H$. Note that $\Delta Q = Q_H - Q_o = Q_H$.

Let P_j represent additional use of resources on treatment (such as multiple and excessive prescriptions) that do not benefit the patient, but bring additional revenues to the doctor and for which

the patient may be willing to pay as long as dispensing a medication is the expected outcome of the doctor-patient interaction. This ‘supplier-induced demand’ (SID) can also represent monopoly pricing (above marginal benefit) for innovative drugs or secret formulations (*mifang*). P_j can range from 0 to P_H . We capture this range by assuming that with probability $p(e_2)$, the doctor succeeds in inducing demand to $P_j = P_H$; otherwise (with probability $1 - p(e_2)$), $P_j = 0$.

In institutional equilibrium, patient and doctor both observe Q_i and P_j . Of course the full benefit (Q_i) and the extent of SID (P_j) will not be evident at the time the doctor writes a prescription or dispenses the medication to the patient. However, in equilibrium with multiple interactions (and the patient’s recourse to self-diagnosis and self-treatment), we assume the patient can eventually observe (average) treatment effectiveness and deduce the extent of SID.

2.3.2 Doctor and patient objectives

Assume the downstream purchaser, the patient, seeks to maximize utility given by perceived benefit from treatment, Q_i , less payment to the doctor.

$$D^K = E[Q_i] - \text{payment} = q(e_1) Q_H - \text{payment}, \quad (1)$$

where K represents the institutional setting: $K = MDD$ or SPD .

To be sustainable, the equilibrium payment under a given institutional setting must also fulfill an ability-to-pay or liquidity constraint. Let the patient’s fraction of the healthcare payment (the co-insurance rate) be θ , where $1 \geq \theta \geq 0$. The maximum patient out-of-pocket spending cannot exceed the patient’s ability to pay, W :

$$\theta (\text{payment}) \leq W. \quad (2)$$

An uninsured patient has $\theta = 1$ and thus faces the tightest constraint. With full or “first dollar” insurance ($\theta = 0$) – very rare historically or in contemporary China – the constraint would never bind even for the lowest wealth levels.⁶

⁶If the population were evenly distributed between maximum and minimum wealth levels, then the higher the co-insurance requirement θ , the larger the fraction of the population for which the ability to pay constraint (2) will bind. We focus on a single representative patient of wealth W , abstracting from financing of the health insurance or health service premiums (presumably tax-based) when $\theta < 1$.

The upstream doctor seeks to maximize revenue less the cost of effort,

$$U^K = \text{Payment} - c(e_1, e_2).$$

Later we will consider a provider putting some weight on patient perceived benefit from treatment, so that provider utility would be $U^K = \text{Payment} + \alpha Q_i - c(e_1, e_2)$. As one might intuitively expect, higher provider ‘benevolence’ or altruism, α , increases quality effort for any given payment, and decreases SID to the extent that SID decreases patient net benefit (increases patient risk from treatment and/or patient spending burden).

2.3.3 The socially efficient benchmark

The socially efficient effort choices maximize social welfare, $S \equiv D^K + U^K$ given by

$$S \equiv q(e_1) Q_H - c(e_1, e_2). \quad (3)$$

In this ‘unproductive multitasking’ setting, the value of the doctor-patient relationship would be highest if the doctor invested only in quality-improvement effort e_1 and never induced demand ($e_2 = 0$). Assume that effort costs take the simple functional form $c(e_1, e_2) = \frac{(e_1)^2 + (e_2)^2}{2}$, and that $q(e_1) = qe_1$ ($q > 0$) and $p(e_2) = pe_2$ ($p > 0$). Then the socially efficient doctor efforts are

$$\begin{aligned} e_1^* &= qQ_H, \text{ and} \\ e_2^* &= 0, \end{aligned} \quad (4)$$

leading to social surplus $S^* = qQ_H^2 (1 - q/2)$. When $q = 1$, $S^* = \frac{Q_H^2}{2}$.

The provider’s actual choices depend on both who owns the Rx “asset” and whether the patient-provider interaction is either (1) relational – with provider commitments to high effort and limited SID rewarded by loyal patient consumers willing to pay a premium for quality – or (2) a “spot market” exchange without hope or commitment to future interactions. The focus is on comparing relational MDD and relational SPD. As BGM highlight, asset ownership shapes renegeing temptations in relational contracts. Whether or not the patient has property rights over where to fill a prescription will affect the provider’s temptation to overprescribe and thus whether MDD will be self-enforcing.

2.4 MD-Dispensing

First consider the provider and patient choices under MDD. This integration of Dx and Rx is analogous to BGM's 'relational outsourcing,' where the downstream (patient) outsources the two desired inputs to health production – Dx and Rx services – to an integrated outside provider organization.

In an ongoing relationship, the purchaser/patient agrees to pay the doctor in three components: R for the contractible minimum quality; bonus b_H if $Q_i = Q_H$; and payment β_H if $P_j = P_H$ (otherwise, $b_i = \beta_j = 0$).

If the doctor believes the patient will follow through with these promised payments, the doctor chooses efforts accordingly:

$$\underset{\langle e_1, e_2 \rangle}{Max} [R + qe_1b_H + pe_2\beta_H - c(e_1, e_2)] \equiv U(e_1^{MDD}, e_2^{MDD}).$$

The equilibrium efforts under relational MDD would then be

$$\begin{aligned} e_1^{MDD} &= qb_H, \\ e_2^{MDD} &= p\beta_H, \end{aligned} \tag{5}$$

yielding doctor payoff $U^{MDD} = R + \frac{q^2b^2 + p^2\beta^2}{2}$. The patient expects a payoff of

$$D(e_1^{MDD}, e_2^{MDD}) = q^2b(Q_H - b) - R - p^2\beta^2.$$

Total surplus under relational MDD is therefore $S^{MDD} = q^2bQ_H - \frac{q^2b^2 + p^2\beta^2}{2}$.

Clearly, the patient will not wish to agree to pay for SID, but may be unable to distinguish P from Q in all cases, and may be willing to pay the doctor a premium or monopoly price ($\beta_H > 0$) to assure quality effort and effective medications in the future. The bonus β_j essentially represents the premium that patients are willing to pay when receiving medication as part of the doctor-patient encounter (i.e., under MDD). We assume that $Q_H \geq b_H > Q_H/2$ and $P_H \geq \beta_H > Q_H/2$ under MDD.

Comparing with (4), we see that the socially efficient levels of effort require $\beta_H = 0$ (which is not possible under MDD unless SID harms the patient and the doctor is altruistic enough) and either $b_H = Q_H$ or sufficient doctor altruism ($\alpha > 0$) to choose e_1^* despite $b_H < Q_H$.

The patient may come to realize that the doctor is charging a monopoly price, or prescribing excessive drugs to increase net revenue, and will be tempted to renege unless the benefits of a continued relationship outweigh the costs of paying P_j . The doctor may be tempted to induce demand as far as possible, but has to weigh the short-term benefit against the loss of reputation and future patronage if the patient goes elsewhere. These temptations to renege on the relational contract depend on what happens when the relationship breaks down.

Assume a one-time doctor-patient interaction under MDD ('spot outsourcing') takes the following form: the patient offers to pay Q_i for Dx and Rx, and the doctor recommends a higher amount P_j consistent with incentives to induce demand. Nash bargaining determines the final payment (amount of Rx and its price): $\frac{Q_i + P_j}{2}$. Anticipating this outcome, the doctor chooses the vector of 'spot MD-dispensing' efforts $e^{SMDD} \equiv (e_1^{SMDD}, e_2^{SMDD})$ according to

$$\underset{\langle e_1, e_2 \rangle}{Max} \left[\frac{1}{2} (qe_1 Q_H + pe_2 P_H) - c(e_1, e_2) \right],$$

so that

$$\begin{aligned} e_1^{SMDD} &= \frac{qQ_H}{2}, \\ e_2^{SMDD} &= \frac{pP_H}{2}, \end{aligned} \tag{6}$$

yielding doctor payoff $U^{SMDD} = \frac{q^2 Q_H^2 + p^2 P_H^2}{8}$. The patient then receives payoff of

$$\begin{aligned} D^{SMDD} &= E \left[\frac{Q_i - P_j}{2} \mid e = e^{SMDD} \right] \\ &= \frac{q^2 Q_H^2 + p^2 P_H^2}{2}. \end{aligned}$$

Spot market MDD thus yields surplus $S^{SMDD} = \frac{3}{8} (q^2 Q_H^2 - p^2 P_H^2)$.

We focus on trigger strategy equilibria: the parties are assumed to honor the relational contract as long as the other party does, and to revert to spot market transactions forever after if the other party reneges. Assuming that r is the discount rate, the doctor will honor the relational contract when the current payment, $b_i + \beta_j$, plus the present value of the future relationship, $\frac{1}{r} U^{MDD}$, exceeds the spot market payment in this period and forever after. That is, the doctor upholds the

relationship if, for all i and j ,

$$\begin{aligned} b_i + \beta_j + \frac{1}{r}U^{MDD} &\geq \frac{1}{2}(Q_i + P_j) + \frac{1}{r}U^{SMDD}, \text{ or} \\ b_i + \beta_j - \frac{1}{2}(Q_i + P_j) &\geq \frac{1}{r}[U^{SMDD} - U^{MDD}]. \end{aligned} \quad (7)$$

The patient will honor the relational contract when the current net benefit ($Q_i - b_i - \beta_j$) and the present value of the future relationship ($\frac{1}{r}D^{MDD}$) exceed the alternative, the net benefit from paying the spot market price ($Q_i - \frac{1}{2}(Q_i + P_j) = \frac{1}{2}(Q_i - P_j)$) and forever after having spot market doctor-patient interactions ($\frac{1}{r}D^{SMDD}$). In other words, the patient renegeing constraint under MDD is

$$\begin{aligned} Q_i - b_i - \beta_j + \frac{1}{r}D^{MDD} &\geq \frac{1}{2}(Q_i - P_j) + \frac{1}{r}D^{SMDD}, \text{ or} \\ b_i + \beta_j - \frac{1}{2}(Q_i + P_j) &\leq \frac{1}{r}(D^{MDD} - D^{SMDD}). \end{aligned} \quad (8)$$

Following BGM's reasoning, if (7) holds for all i and j , then it must hold for the smallest value of $b_i + \beta_j - \frac{1}{2}(Q_i + P_j)$, which is 0. Moreover, if (8) holds for all i and j , then it must hold for the largest value of $b_i + \beta_j - \frac{1}{2}(Q_i + P_j)$, which arises when $b_H = Q_H$ and $\beta_H = P_H$, so that $b_i + \beta_j - \frac{1}{2}(Q_i + P_j) = \frac{1}{2}(Q_H + P_H)$. These two conditions can be combined to form the following necessary condition for MDD to be self-enforcing:

$$\frac{1}{2}(Q_H + P_H) \leq \frac{1}{r}(S^{MDD} - S^{SMDD}). \quad (9)$$

where S^K denotes the total surplus achieved under each contracting arrangement, $K \in \{MDD, SMDD\}$. Notice that MDD is an equilibrium only when the maximum temptation and ability to induce demand, P_H , is not too large.

To be sustainable, MDD must also fulfill the ability-to-pay or liquidity constraint, (2). The maximum patient out-of-pocket spending under the MDD doctor-patient relationship ($\max(b_i + \beta_j)$) and under the spot market alternative ($\frac{Q_H + P_H}{2}$) cannot exceed the patient's ability to pay, W : $\theta(Q_i + P_j) \leq W$, or

$$P_H \leq \frac{W}{\theta} - Q_H. \quad (10)$$

Here again we see that MDD can be self-enforcing only if the maximum magnitude of induced

demand and quality Q is not too high. Each dollar spent on SID detracts from ability to pay for quality effort (and vice versa). When patients lack insurance coverage ($\theta = 1$), ability to pay is likely to be the binding constraint on SID, with P_H defined to be $P^H(\theta)$ by (10).

A decrease in θ – an increase in coverage generosity, such as adopting social health insurance – enables patients to buy more health care, relaxing the ability to pay constraint on both quality (classic moral hazard) and SID (supply-side moral hazard): $\frac{dP^H}{d\theta} < 0$. Some evidence in support of this assumption comes from Iizuka (2007), who finds that Japanese dispensing physicians will forego one dollar in mark-up in exchange for a 28 cent reduction in patient out-of-pocket cost. He hypothesizes that patient ability to discipline the provider, by referrals or control of repeat visits (and thus future demand and revenue for the doctor) constrains SID. Increasing generosity of coverage weakens this demand-side constraint.

I also assume that doctors’ maximum ability to induce demand depends on available technology, $TECH$.

Claim 2 $P^H(\theta; TECH)$ with $\frac{\partial P^H}{\partial TECH} > 0$: *As the capabilities of medicine grow, the social opportunity costs from SID also grow.*

As a result, technological change will play a role in determining when MDD and SPD are self-enforcing institutions.

2.5 Separation of Prescribing and Dispensing

If the downstream purchaser “owns the prescription asset” and fills the prescription at an independent pharmacy, then the renegeing temptation differs, although the actions and outcomes under a relational SPD contract can achieve the same outcomes as under MDD (see BGM result 1, p.64).

Under SPD, the patient is relatively “empowered” vis-a-vis the provider. (See Hart (1995) on how asset ownership under incomplete contracting conveys power.) The doctor has no incentive to induce demand because all dispensing revenues flow to independent pharmacies. The doctor also has limited ability to charge a monopoly price for a *mifang*, since the patient (and perhaps pharmacist) can see the prescription and how it is matched to symptoms and “defect” from the doctor-patient relationship to engage in self-diagnosis or pharmacist ‘quasi-prescribing’ in the future. The doctor may also have little incentive to invest in quality-improvement effort. In addition, compared to MDD, the patient under SPD must incur higher travel costs to visit both a doctor and a separate

pharmacy. When will it be socially efficient to require patients to pay these extra travel costs and to invest in the contracting institutions to uphold such a separation equilibrium? In other words, when will SPD facilitate doctor-patient relationships that generate social surplus sufficient to justify abolishing MDD?

Under SPD, the patient ‘employs’ the doctor as his or her agent to diagnose the medical condition and prescribe a remedy. Under ‘relational employment,’ the upstream provider’s temptation is not inducing demand but rather stinting on effort. Indeed, when SPD is enforced, the fall-back spot market transaction is that the doctor provides only minimum contractible quality. Under a one-time doctor-patient interaction with independent pharmacies, the patient utility is $D^{SSPD} \equiv Q_o - R$ and the doctor utility is $U^{SSPD} = R$; the resulting social surplus is $S^{SSPD} = Q_o - R + R = Q_o = 0$.

Since the doctor no longer reaps benefits from Rx, the patient need not promise to pay for P_j : $\beta_H = 0$ and therefore $e_2^{SPD} = 0$. Thus, by design, SPD removes the opportunity for excessive prescriptions motivated by prescriber-dispenser profits (P_H). Doctor utility becomes $U^{SPD} = R + qe_1b_H - c(e)$, yielding quality effort $e_1^{SPD} = qb_H$. Thus $U^{SPD} = R + \frac{(qb_H)^2}{2}$.

The downstream patient/purchaser must pay for the contracting costs of legislating and enforcing separation (each period), Ω , as well as bear the additional travel or transaction costs associated with separate doctor offices and pharmacies, T_t . Patient utility under SPD relational contracting is thus

$$D^{SPD} \equiv E [Q_i - R - b_i \mid e_1 = e_1^{SPD}] - \Omega - T_t.$$

$T_t + \Omega$ is the per-period cost of a self-enforcing relationship under SPD (which need not be paid again if the relationship breaks down and the parties revert to spot market interactions). The patient expects a payoff under relational SPD of

$$q^2b_H(Q_H - b_H) - R - \Omega - T_t \equiv D^{SPD}.$$

The patient can choose to honor the contract, paying b_i and receiving $\frac{1}{r}D^{SPD}$ in the future. Alternatively, the patient can renege, which would save the current bonus payment but imply no doctor effort above the minimum contractible level in the future. The patient honors the relational contract if the maximum promised bonus is less than the incremental benefit of relational contracting compared to spot market contracting. From the patient’s perspective, then, SPD can

be self-enforcing when $-b_i + \frac{1}{r}D^{SPD} \geq \frac{1}{r}D^{SSPD}$, which can be re-written as $\frac{1}{r}(D^{SPD} - D^{SSPD}) \geq \max(b_i) = b_H$, or

$$\frac{1}{r}(qe_1^{SPD}Q_H - T_t - \Omega) \geq b_H. \quad (11)$$

The doctor can choose to honor the relational contract, accepting $b_i = b_H$ when $Q_i = Q_H$ and $b_i = 0$ when $Q_i = Q_0$, and expecting $\frac{1}{r}U^{SPD}$ in the future. Or the doctor can refuse the bonus as too low for the requisite effort, and instead revert to one-time interactions with patients for prepayment R and minimum contractible effort. The doctor will honor the relational contract if the incremental value of continuing the relationship exceeds the minimum expected “pay for performance” bonus: $b_i + \frac{1}{r}U^{SPD} \geq \frac{1}{r}U^{SSPD}$, or

$$\min(b_i) = 0 \geq \frac{1}{r}(U^{SSPD} - U^{SPD}). \quad (12)$$

Combining these two renegeing constraints as in BMG, we find that the institution of separating prescribing from dispensing is self-enforcing if the relational (noncontractible) bonus payment b_H is not too great (relative to the payment for contractual minimum quality):

$$\begin{aligned} \max b_i - \min b_i &= b_H \leq \frac{1}{r}(D^{SPD} - D^{SSPD}) - \frac{1}{r}(U^{SSPD} - U^{SPD}), \text{ or} \\ b_H &\leq \frac{1}{r}(S^{SPD} - S^{SSPD}) = \frac{1}{r}S^{SPD}. \end{aligned}$$

Notice that unlike for MDD, ability to induce demand (P) plays no role in the renegeing constraint for SPD.

SPD must also fulfill the ability to pay constraint (2), but SPD by definition removes P^H from the patient out-of-pocket burden and thus may relax the constraint. In effect SPD substitutes a payment to enforcement institutions ($\Omega + T_t$) for the extra payment to the provider (P), allowing patient payments at time of service use to be better targeted on increasing the value of treatment, Q .

Using the simplified functional forms,

$$S^{SPD} = q^2bQ - \frac{q^2b^2}{2} - \Omega - T_t,$$

which can be written as $bQ - \frac{b^2}{2} - \Omega - T_t$ if $q = 1$. Thus the renegeing constraint for SPD is fulfilled

when $b_H \leq \frac{1}{r} \left(bQ - \frac{b^2}{2} - \Omega - T_t \right)$, or r is small enough to support reputational enforcement of SPD.

2.6 MD-dispensing and separation reforms

What does this simple theory suggest about the determinants of MDD or SPD as self-enforcing social institutions governing the doctor-patient relationship?

The self-enforcing institution of MDD dominates the self-enforcing institution of SPD if and only if $bQ - \frac{(b^2 + \beta^2)}{2} > bQ - \frac{b^2}{2} - \Omega - T_t$, or $\frac{\beta^2}{2} < \Omega + T_t$. Assuming $\max(\beta_H) = P^H$, then MDD dominates SPD when

$$\frac{(P^H)^2}{2} < \Omega + T_t. \quad (13)$$

Proposition 3 *The higher the contracting and perceived travel costs of SPD, the wider the range of parameters over which MDD is self-enforcing.*

Recall the earlier claim that T_t increases in the number of periods a society has been under MDD: generations of patients associate receipt of a medication with a successful doctor-patient visit, and therefore perceive there to be a high cost of *not* receiving a drug from a doctor. In this way, MDD becomes a self-reinforcing institution (Greif 2006). Each individual does not foresee this reinforcing of MDD, however. If players are unaware of the reinforcement mechanism, then at time τ they imagine that the costs of switching to SPD will remain T_τ into the future, although in fact it will increase if MDD continues. Thus T_t is not a fixed parameter, nor an endogenous variable; it is a ‘quasi-parameter’ (Greif 2006).

Proposition 4 *The longer a society remains under MDD, the higher the associated costs of SID (P^H) can be before implementing SPD becomes the efficient self-enforcing social institution.*

Institutional entrepreneurs may come to understand that the MDD status quo is self-reinforcing and propose reforms that preserve as many of the institutional elements of MDD as possible while still moving toward SPD. Many of China’s payment reforms can be interpreted in this light. For example, reforms to separate drug revenues from hospital expenditures take place without any requirement that patients change their expectations or behavior: patients still receive medicines at the hospital where they are diagnosed and treated. The reform attempts to de-couple provider income and drug utilization while nothing changes for patients.

Recall that the potential for SID in any given period – P^H – is also itself a function of insurance and technology, and is thus a second quasi-parameter shaping the historical evolution of doctor-patient institutions as either MDD and SPD, which we will consider in more detail below.

2.6.1 Variation with r and P_H

Figure II in BGM illustrates how the ‘efficient organizational form’ – MDD (‘relational outsourcing’) and SPD (‘relational employment’) – varies with the discount rate r and magnitude of P_H (ΔP). For high discount rates, neither patient nor physician cares much about the future, and therefore reputation-based relational contracts are not self-enforcing; spot market interactions prevail. At very low discount rates, both institutional arrangements can achieve socially efficient outcomes. Probably most interesting and policy-relevant is the intermediate range of low to moderate discount rates. In this range of r , the purchaser and provider agree to cooperate in an ongoing relationship, with MDD dominating for low P_H and SPD dominating for high P_H .

BGM’s “first result” is that switching from relational outsourcing to relational employment is an efficient response to widely varying supply prices (BGM, p.64). In the present context, this result implies the following:

Proposition 5 *Separation policies – switching from MDD to SPD – can be an efficient response to increases in P_H , the potential loss in social surplus from maximum supplier-induced demand.*

2.6.2 Variation with T_t , Ω , and P_H

Let us focus on the range of moderate discount rates, and examine how the socially sustainable arrangement varies according to contracting and travel costs as well as the magnitude of potential supplier-induced demand. Figure 1 illustrates how the quasi-parameters T_t , Ω , and P_H impact the range over which MDD and SPD are self-enforcing institutions, assuming an intermediate range of discount rates that can support relational contracting. I also simplify notation by assuming $q = p = 1$.

The horizontal axis once again graphs P_H . The vertical axis shows the costs associated with enforcing SPD, $T_t + \Omega$. The upper right-hand side, with high $T_t + \Omega$ and low P_H , represents a range of over which MDD is the efficient equilibrium institution. The lower left-hand side, with low $T_t + \Omega$ and high P_H , illustrates the range of over which SPD is the efficient self-enforcing institution. The dividing curve illustrates when $S^{MDD} > S^{SPD}$, as given by $\frac{(P_H)^2}{2} < T_t + \Omega$.

What determines the magnitude of P_H ? Technological change continuously pushes up the capabilities of medicine, and with it, the potential welfare losses associated with SID. Thus, all else equal, SPD is more likely to prevail as doctor recommendations influence utilization of a wide range of expensive health care technologies.

A second critical determinant of P_H is the generosity of social or private health insurance or healthcare coverage (such as through National Health Insurance or a National Health Service). Adoption of universal coverage is plausibly associated with a significant one-time increase in P_H : reducing the patient burden of out-of-pocket spending will relax the ability-to-pay constraint on SID.

Indeed, I would posit a political mechanism that links increases in $TECH$ to reductions in patient out-of-pocket burden (θ). The increasing capabilities of medicine trigger institutional change toward universal coverage because as $TECH$ increases, access to healthcare becomes more unequal and has more significant impacts on life chances. Given a certain level of social solidarity, or political threshold of acceptance for disparities in access to healthcare technology, increases in technology that increase spending (and thus strain the poor's ability to pay) will trigger political reform to adopt or extend risk pooling for medical expenses (decreasing θ). To the extent that per capita income is correlated with the level of healthcare technology perceived as available in a society, the pattern of higher public spending at higher per capita income levels is consistent with this hypothesis. Further empirical evidence consistent with this hypothesis comes from the acceleration of adoption of universal coverage reforms at ever lower levels of per capita income.⁷ Of course there also is evidence that insurance coverage can spur technological change (e.g. Finkelstein 2007).

I summarize this discussion with the following proposition:

Proposition 6 *Technological change (increasing $TECH$) directly increases P^H (because technology tends to increase spending) and indirectly increases P^H through an induced increase in coverage:*

$$\frac{dP^H}{dTECH} = \frac{\partial P^H}{\partial TECH} + \frac{\partial P^H}{\partial \theta} \frac{\partial \theta}{\partial TECH} > 0.$$

(+)(-)(-)

(see(10)). Therefore separation policies are more likely when there has been (i) rapid technological

⁷Japan achieved universal healthcare coverage in 1961, when GDP per capita (in constant 2000 US\$) was 7883. South Korea achieved universal coverage in 1989 with GDP per capita of 6130. Thailand adopted universal coverage in 2003 at much lower per capita GDP, and China may be on track to achieve universal coverage (albeit at a low benefit level) by 2010, although current per capita GDP is only 1595. (All GDP figures are in constant 2000 US dollars, relying on the World Bank WDI online database; see Eggleston 2008c.)

change in healthcare and/or (ii) a large increase in generosity of insurance coverage, which may itself be a political result of technological change increasing disparities of access beyond the acceptable social threshold.

How do these propositions help to explain the complex of incentives in China's healthcare financing and delivery system? Societies would be arrayed in Figure 1 in different ways, with many European countries toward the southwest (early separation) and the US and UK towards the middle (later separation).

I argue that China historically had high contracting costs and low potential for SID. MDD was a self-enforcing and self-reinforcing institution. Over a long period of time, China moved from southwest to northeast on Figure I: contracting costs have declined somewhat with development, but MDD was self-reinforcing as patients and doctors came to believe that dispensing was the natural outcome of any effective doctor-patient encounter. On net, arguably $\Omega + T_t$ increased (until recently). Technological change has pushed P_H ever higher, but self-reinforcing MDD increased T_t ; together they pushed China upward and left, perhaps close to but only recently crossing over the boundary between MDD and SPD as self-enforcing institutions.

Thus it is ahistorical to think that China's embedding of MDD in its healthcare system simply reflects backwardness compared to Europe, which separated prescribing from dispensing much earlier. The institutional context of the two has been quite different. China's neighbors sharing MDD histories have only recently implemented SPD (Japan is still 'in progress'; see Iizuka 2008 and Tomita 2008).

The proposition above predicts that China will adopt more rigorous separation policies as it commits to universal coverage and (gradually) replaces demand-side constraints with supply-side constraints on spending. Indeed, China's East Asian neighbors provide some evidence in support of this hypothesis. Separation policies have been enacted in both Korea and Taiwan soon after adopting NHI. South Korea achieved universal coverage under NHI in 1989, and implemented a strict (and very politically contentious) separation of prescribing from dispensing in 2000. Taiwan implemented NHI in 1995, and began only three years later with separation reforms, albeit with less success (since doctors are allowed to hire their own pharmacists and hospitals can retain their outpatient pharmacies). China has been experimenting with SPD at the same time that it is expanding insurance coverage. Conflicts over the right to prescribe and dispense may escalate even as researchers and policymakers increasingly question the appropriateness of the prevailing MDD

system for aligning health care delivery with the social goals of quality care at reasonable cost.

2.6.3 Heterogeneity

Up to this point, I have simplified the analysis by assuming a single institutional structure was self-enforcing for all patients and providers in a given society. Clearly, however, the benefits and costs of MDD can differ across population sub-groups.

For example, the theory predicts that MDD will be the more efficient institutional arrangement when the perceived costs associated with traveling to two different locations are sufficiently high. This arguably explains why hospitals dispense medications to inpatients, even when they are prohibited from dispensing to outpatients (as in Korea). Another, somewhat related case arises when the physician diagnosis and treatment services are inextricably linked to the dispensing of medication, such as for injections ostensibly requiring physician oversight. Theory predicts that the hazard that this arrangement must confront is the temptation for SID on profitable drugs. Indeed, just such a controversy plagues injectable drugs in the US; see for example the editorial in the New York Times in May 2007 which argues that “federal laws already bar drug companies from paying doctors to prescribe medicines in pill form. That prohibition should be extended to injected and intravenous medicines.” [NYT editorial, “The Danger in Drug Kickbacks” May 14, 2007, <http://www.nytimes.com/2007/05/14/opinion/14mon1.html>].⁸

Probably the most common exception to SPD arises when separation is prohibitively costly because of insufficient human resources relative to high travel costs, such as in rural areas of even high-income countries with long traditions of separation (e.g. France or the UK). Ample historical evidence suggests that occupational distinctions between physicians and pharmacists have been blurred when one or the other of the professions was scarce relative to effective demand. This partial MDD usually took the form of pharmacists practicing medicine when physicians were not available (such as a royal ordinance in France in 1724 [Sonnedecker 1976, p.71] or in rural NHS or many developing countries today). In some cases physicians were licensed to dispense when pharmacies were scarce. In South Africa in the 1980s, the government began granting physicians licenses to dispense medicines to serve communities without pharmacies; the access to licenses became such that by the late 1990s, four out of five doctors are permitted to dispense medications

⁸FTC studies have also alleged that US optometrists regularly prescribe unnecessary treatment (Wolinsky 1993, p.380).

(Gilbert 2001, p.102). India is another example of a developing country where restricting access to ‘less than fully qualified providers’ would deprive much of the population of access to therapies commonly used and considered effective.

A final important dimension of heterogeneity involves the distinction between acute care episodes and management of chronic disease. When a patient has a chronic disease, the travel costs associated with frequent adjustment of prescriptions may also weigh in favor of integrating some diagnosis and dispensing functions. If a patient requires frequent interaction with providers for managing illness, then integrating some part of the evaluation and management with treatment at locations more numerous throughout the community (such as pharmacies) reduces the total cost of managing that patient, perhaps with little or no reduction in quality. As China’s burden of disease from chronic noncommunicable diseases continues to grow, it is worth considering to what extent preserving a limited amount of prescriber-dispensing – such as allowing qualified pharmacists to adjust prescriptions for diabetic patients or coach on self-management – might be warranted.

3 Price regulation and payment reforms

This section examines the conceptual underpinnings of China’s regulated prices and hospital payment reforms, especially the current widespread experimentation with case-based payment, global budgets, and ‘separation of revenues and expenditures.’

3.1 A simple model

Consider a provider (physician, clinic, or hospital) providing multiple services to patients. Let m_j represent the spending on health service j , and e_j denote provider effort on service j . Patient benefits $v_j(m_j, e_j)$ are increasing and concave in spending and effort. In this model, the cost of SID is associated with the provider choice of excessive treatment spending ($m > m^{**}$) rather than a separate effort e_2 as in the MDD model above.

A patient’s total benefit from healthcare is the sum of benefits from each health service: $v(m, e) = \sum_j v_j$. Recall that (1) defines the downstream consumer’s utility as $D = E[Q] - \text{payment}$. Now, $E[Q] = \sum_j v_j$ and $\text{payment} = \sum_j \theta_j m_j$, so that patient utility is $D = \sum_j (v_j - \theta_j m_j)$.

Assume that provider effort e_j on service j increases patient utility $v_j(m_j, e_j)$ for a given level of spending, at a decreasing rate: $\frac{\partial v_j(m_j, e_j)}{\partial e_j} > 0$; $\frac{\partial^2 v_j(m_j, e_j)}{\partial e_j^2} < 0$. Assume quality effort increases the

marginal benefit of treatment spending for each service: $\frac{\partial^2 v_j(m_j, e_j)}{\partial e_j \partial m_j} > 0 \forall j$. In other words, effort and spending are complements.⁹ Denoting partial derivatives with a subscript, in a one-service model, $\frac{\partial v(m, e)}{\partial e} = v_e > 0$.

Quality effort e causes provider disutility $c(e)$, increasing and convex in effort.¹⁰ Purchasers may pay a clinic or hospital a ‘quality bonus’ for specific quality improvements. Assume the patient pays the doctor b_j per unit of e_j . The payment may be contracted (such as ‘pay for performance’ for doctors in higher-accreditation-level hospitals) or relational (based on noncontractible efforts, as in the above MDD-SPD model).

Recall that provider utility was defined as $U = \text{Payment} + \alpha Q_i - c(e_1, e_2)$. With insurance, demand-side and supply-side cost sharing can be set separately so that the *payment* from patient to provider at point of use does not necessarily equal the total *payment* received by the provider for that service. Let provider net revenue π per patient consist of three components: pre-payment R ; pay-for-performance incentives b_j for e_j ; and supply-side cost-sharing for each service, s_j , defined by the fraction of spending $(1 - s_j)$ that the payer reimburses:

$$\pi = R + \sum_j [b_j e_j + (1 - s_j)m_j - m_j] = R + \sum_j [b_j e_j - s_j m_j]. \quad (14)$$

With this definition of payment, and assuming that providers care about patient benefit ($\alpha > 0$), provider utility becomes

$$U = \alpha v + \pi - c(e). \quad (15)$$

The payer regulates the price for each service and thus implicitly specifies supply-side cost sharing for each service, s_j . In China, to provide implicit insurance, administered prices were set so that basic non-invasive services were cheap, with prices that often did not cover even marginal cost, implying supply-side cost sharing ($s_j > 0$). Services that the doctor provided but could not charge patients or insurers, including public health measures not explicitly reimbursed, faced full supply-side cost sharing ($s_j = 1$). Other services – particularly drugs and high tech diagnostics or procedures – enjoyed more generous prices implying a positive profit margin ($s < 0$). For example, hospitals are officially allowed a 15 percent mark-up for dispensing drugs, so the patient or insurer

⁹ At several places in the analysis I discuss how results may differ if spending and effort were substitutes instead.

¹⁰ In a multi-service model (see Eggleston 2005), increasing one effort may increase the marginal cost of other efforts: $\frac{\partial^2 c}{\partial e_j \partial e_k} > 0$.

pays $1.15m$; the hospital net revenue is $1.15m - m = 0.15m$, or $s = -0.15$ for dispensing drugs. Note that this margin for drugs was built into the incentive system very early, in the 1950s (Zhang Qiong, 2008), and continues to today. Arguably this margin did not significantly impact behavior under central planning when the majority of provider revenues came from lump-sum budgets (through R in our model) rather than on a fee-for-service basis. After the 1980 reforms, however, prospective budgets declined significantly and regulated prices – with their implied differences in profitability and supply-side cost sharing – became significant determinants of provider behavior.

For simplicity, consider a single-service model. The efficiency benchmark, achieved if the purchaser could directly contract on quality and spending, would be

$$\underset{\langle m, e \rangle}{Max} [v(m, e) - m - c(e)], \quad (16)$$

implying

$$\begin{aligned} v_m(m^{**}, e^{**}) &= 1, \\ v_e(m^{**}, e^{**}) &= c'(e^{**}), \end{aligned} \quad (17)$$

where m^{**} and e^{**} denote the efficient (first-best) levels of spending and effort.

The provider can be said to induce demand for a service ($P_j > 0$) when recommending spending on that service in excess of the amount that equates marginal benefit with marginal cost (the amount that a fully informed self-paying patient would desire): $m > m^{**}$. (I use this definition since MDD prevailed when patients were uninsured; when patients have insurance, of course, patient moral hazard would imply spending in excess of the efficient amount, and SID would properly only be that amount in excess of what the insured patient desires: $SID = m - m^*$ where $v_m(m^*, e^*) = \theta$. Note that if the patient were fully insured, SID implies spending so much that there is negative marginal benefit, $v_m < 0$. The social cost of such ‘flat of the curve medicine’ (P in the above model) includes both the reduced patient benefit ($v(m^{**}) - v(m)$) and the excessive resource use ($m - m^{**}$).

When the payer cannot specify in a contract the appropriate spending and effort for each patient, the provider chooses m and e to maximize utility (15), responding to bonus and cost

sharing incentives. The provider's maximization program is

$$\underset{\langle m, e \rangle}{Max} [\alpha v(m, e) + R - sm + be - c(e)]. \quad (18)$$

This provider utility function is assumed to be strictly concave. The first-order conditions define spending $m^*(\alpha, s, b)$ and effort $e^*(\alpha, s, b)$ as functions of provider benevolence and the payment parameters:

$$\begin{aligned} v_m(m^*, e^*) &= \frac{s}{\alpha}, \\ \alpha v_e(m^*, e^*) + b &= c'(e^*). \end{aligned} \quad (19)$$

Comparison to the efficiency benchmarks shows several deviations from first-best. For one thing, the provider may not be a perfect agent for the patient: $\alpha \neq 1$. Consider a modestly benevolent provider, $\alpha < 1$. (The previous MDD model represented the extreme, since the altruism parameter was zero; the doctor only cared about payment.) Since in this case $\alpha v_m < v_m$ and $\alpha v_e < v_e$, the provider prefers to stint on spending and effort unless explicitly compensated for the resource and utility losses associated with m and e , respectively.

Appropriate payment, through choice of s and b , can help to compensate for the distortions arising from provider imperfect agency. Generous administered prices translate into low or even negative supply-side cost sharing, s , which in turn lowers the provider's marginal cost of spending and increases m : $\frac{\partial m}{\partial s} = \frac{\alpha v_{ee} - c''}{H} < 0$. And of course a positive bonus b for effort raises the provider's marginal benefit from quality effort, which raises quality¹¹:

$$\frac{\partial v}{\partial b} = \frac{\partial v}{\partial e} \frac{\partial e}{\partial b} = -\frac{\alpha v_e v_{mm}}{H} > 0. \quad (20)$$

When large contracting costs Ω preclude contracting on quality, payer bonuses b are limited to relational payments (reputational effects). In fact the model suggests that this noncontractability is one reason why informal under-the-table payments (*hongbao*) are so prevalent in China.

The inability of an insurer or patient to contract on e does not preclude bonuses b altogether, however, since a large fraction of China's doctors are employees of hospitals in both urban and rural areas. Thus the doctors' incentives depend upon the 'pass-through' of hospital payments

¹¹The numerator is negative (because of diminishing marginal returns to spending, $v_{mm} < 0$) and the denominator $H = [\alpha v_{mm} (\alpha v_{ee} - c'') - (\alpha v_{me})^2]$ is positive by strict concavity of U .

to the individual doctors, as captured by their bonuses and promotional incentives. A hospital administrator seeking to maximize net revenue would want to set bonuses to reflect the supply-side cost sharing in hospital payments, such as bonuses for high volume under FFS, or bonuses directly linked to hospital profits. Indeed, both are common in China, as discussed by the empirical studies of Liu and Mills (2003) and Eggleston et al. (2008b).

A hospital administrator seeking to maximize net revenues would want to ‘pass through’ the hospital’s supply-side cost sharing s_j to the hospital’s doctors; as Eggleston and Yip (2005) show, this could be done by setting the formula for doctor bonuses such that their supply-side cost sharing, S_j , is strongly and positively correlated with that facing the hospital, s_j (e.g., $\frac{s}{\alpha} = S$).

With e noncontractible and quality bonuses b thus limited, the underlying payment scheme – s_j – plays the dominant role in shaping provider behavior. Generous FFS margins ($s < 0$) induce excessive spending on the profitable services, pharmaceuticals and high tech diagnostics. It is not surprising that the empirical literature on quality focuses primarily on the propensity for over-prescribing and overprovision associated with profitable services, and exacerbated by the relaxation of the ability to pay constraint for insured consumers. Thus reports of excessive prescribing, overuse of antibiotics and other antimicrobials, hospital purchasing of drugs according to the hospital’s margin, patient self-treatment to avoid SID, and so on, pervade the healthcare analyses in China (e.g. see reviews for overall service delivery and pharmaceuticals, respectively, in Eggleston et al. 2008 and Sun et al. 2008).

Patients are far from ignorant about these incentives and resulting provider behavior. In fact, patients widely expect providers to induce demand and complain about excessive treatments. Evidence for this perception comes from, among other factors, the phenomenon of Chinese hospitals voluntarily adopting fixed case-based payment as a way to attract patients.

3.2 Hospital self-adoption of case-based payment

The most famous and pioneering case of a hospital announcing a self-imposed price ceiling per case is that of Jining Medical College hospital in Shandong province, which since April 2004 adopted price ceilings for 67 diseases, now expanded to 128 specific diseases (Wu Guanghua 2007; Meng 2008). For patients with these diagnoses, hospital representatives and the patients discuss the fee and care protocol and sign an agreement prior to treatment.

That adopting a maximum payment per case would attract patients when they perceive SID

as the predominant problem is simple to show in this model. Assume the number of patients is given by $n(m)$. Demand increases in spending if patients suspect stinting, such as under capitation or managed care ($\frac{dn}{dm} > 0$ if patients believe $m < m^{**}$). Demand decreases in spending if patients suspect SID ($\frac{dn}{dm} < 0$ if patients believe $m \gg m^{**}$). Announcing a price ceiling is analogous to announcing a switch from $s < 0$ to $s > 0$; since $\frac{\partial m}{\partial s} < 0$, the switch implies a significant decrease in spending. This announcement only serves to attract patients if $\frac{dn}{dm} < 0$, that is, if patients perceive SID to be the primary concern.

This model predicts that hospital self-adoption of case payment is most likely for (i) ‘middle-level’ hospitals; (ii) services for which patients can plan ahead, such as childbirth and elective surgery; and (iii) when the hospital administration is willing and capable of changing the way they reward their hospital staff, despite the external payment system for the hospital remaining FFS. Preliminary evidence appears to support these hypotheses. Jining hospital is not one of the largest tertiary hospitals in Beijing or Shanghai, but it is affiliated with a medical college and thus is not one of the many small county or urban primary care hospitals. The large urban academic medical centers in China have long queues for services every day, and thus face no shortage of demand that might spur such a reform. The smallest hospitals, such as township health centers or private for-profit entrants in urban areas, do suffer from low occupancy rates and would like to attract patients through such a tactic. But they lack the quality reputation to attract patients based on low price alone, since for them the quality-stinting concern would arise more prominently. A ‘middle-level’ hospital has much to gain by attracting more patients and can do so with a case-payment promise because the overwhelming concern is SID and the hospital’s reputation provides some guarantee against the extreme stinting that might otherwise occur. Hypothesis (iii) is supported by the fact that Director Wu of Jining hospital himself stresses the importance of changes in personnel policy – competition for the position of clinical department head based on quality assurance targets and reductions in drug utilization – that he had to initiate to implement the case payment reform (see Meng 2008 for summary of interview).

Thus, a simple and intuitive model of provider payment incentives can elucidate numerous facets of China’s current payment reforms, including specific empirically testable hypotheses about how, when, and where reforms will take hold. We can further enrich this simple model by assuming that spending and effort are complements (or substitutes) and by allowing for provider multitasking across multiple dimensions of effort for each service.

3.3 Extensions: Complements and Multitasking

When spending is complementary to quality effort ($v_{me} > 0$), increasing pay-for-performance increases both quality and spending. By contrast, increasing supply-side cost sharing promotes cost control at the expense of complementary effort:

Supply-side cost sharing (s)	Pay-for-performance (b)
$\frac{\partial m}{\partial s} = \frac{\alpha v_{ee} - c''}{H} < 0$	$\frac{\partial m}{\partial b} = \frac{\alpha v_{me}}{H} > 0$
$\frac{\partial e}{\partial s} = \frac{-\alpha v_{me}}{H} < 0$	$\frac{\partial e}{\partial b} = \frac{-\alpha v_{mm}}{H} > 0$

Thus, the payer can promote quality effort with high-powered pay-for-performance or with low-powered supply-side cost sharing (mixed payment).

The comparative statics on provider ‘benevolence’ or altruism (not shown) suggest that when the marginal benefit of spending is positive, provider benevolence reduces stinting on both treatment and quality effort. Better agents for patients are less responsive to payment incentives when those incentives conflict with patient interests.

Of course, good agents also indulge patient moral hazard, which is financially rewarded under generous FFS payment such as the margin for drugs, so provider agency does not necessarily imply social efficiency. This is one reason why SPD did not reduce pharmaceutical spending much in Korea (indeed, prescription drug spending increased relative to trend in other economies; see Figure 2) since providers tended to prescribe brand drugs as a signal of quality and ignored costs to the insurer. However, blaming provider ‘greed’ and ‘lack of professionalism’ smacks of the old socialist mindset – pursuit of self interest is evil – and absolves the policymakers of culpability for designing a payment system that financially rewards (and attracts to the profession) those who are most willing to induce demand for profitable services.

Matters are further complicated by the reality that doctors and hospitals routinely provide multiple services, so that multitasking across contractible and noncontractible dimensions of quality is a concern. Eggleston (2005) uses a two-service model to show that mixed payment helps pay-for-performance avoid perversely rewarding providers for skimping on noncontractible dimensions of quality.

For example, several Chinese payers (such as municipal social insurance schemes) have experimented with case-based payment for select medical conditions (Eggleston et al. 2008; Meng 2008). As noted above, switching from FFS to fixed payment per patient, as in case-based payment reform, reverses the incentive for SID on the services included in the case payment. Indeed, the

primary concern arising from the case payment reforms in China has been evidence of stinting, such as early discharge, refusing costly patients or readmissions (e.g., Zhang 2008; Meng 2008). The model here, extended to the case of multitasking across multiple services (Eggleston 2005), suggests intuitively that providers would stint less if service quality were contractible and rewarded under pay-for-performance, b_j ; but also and more generally (given limitations on contractability, $\Omega > 0$), that providers would stint less if the case payment were instead mixed payment ($0 < s < 1$, i.e., the marginal payment is greater than zero).

3.4 Dynamic incentives: Ratchet effect and soft budget constraints

Finally, although the one-service model of payment incentives presents a static picture, in reality the payment system and ownership structure has important dynamic elements as well, with roots in China's centrally planned past. For example, reforms toward (negotiable) global budgets, a common reform in China's urban areas, may appear to give strong incentives for cost control in any given year; but the payer's ability to set the budget for the next year based on the previous year's cost saving achievements – by, for example, not updating at the same pace as inflation – builds in a ratchet effect (Eggleston and Yip 2001) similar to those infamous under quota systems: over-fulfilling the quota may lead to rewards this year, but also tends to lead to an even more ambitious target for next year and thus a “ratcheting up” of performance standards.

Another example is that of soft budget constraints (Kornai 1986; Kornai, Maskin and Roland 2003). For example, Eggleston, Shen, and colleagues (2008b) investigate the financial support that hospitals receive from government and other supporting organizations using two waves of data from over 300 hospitals in Guangdong province. They find that government hospitals making larger financial losses in 2002 are the most likely to receive government support in 2004. Dividing the hospitals according to their quintile of the 2002 profit margin distribution, the authors find that the probability of receiving government financial support in 2004 goes down gradually as hospitals move up the profit margin categories in 2002, a finding that is consistent with the soft budget constraint theory. None of the for-profit hospitals receive any government financial support.

This quantitative study and other anecdotal evidence suggests that soft budget constraints continue to shape the behavior of government-owned hospitals in China, despite their heavy reliance on patient out-of-pocket payments. Moreover, there appears to be an interesting and deleterious mixture of soft budget constraints (for key government hospitals) and overly hard budget constraints

(e.g., for local government insurers who are afraid if they reimburse many patients they will go bankrupt, without any bail-out). Predictably this situation leads to tensions, with patients caught in the middle, receiving treatment in inefficient hospitals but unable to receive reimbursements for expensive services because insurers are vague *ex ante* about what can be reimbursed (benefit entitlements), while they are stringent with spending *ex post*. The latter situation is improving, as it became a social problem and the government gave orders that local insurers and NCMS schemes should not accumulate surpluses.

There also is an interesting issue with payment reforms related to hard budget constraints: providers do not want to accept the risk of treating costly patients who would spend more than the fixed payment per patient, so they refuse treatment (rather than accepting the risk of losing money on some patients but gaining money on other patients). In this case the provider is mistakenly interpreting the fixed prepayment as a hard budget constraint at the individual patient level, and transferring the burden back to the patient. This situation predictably will change with improvements in transparency and legal recourse for patients in China, but argues once again in favor of mixed payment.

4 Conclusion

This paper focuses on positive analysis of why China's health system incentives evolved the way they did. The first section analyzes the institution of physician dispensing historically up through the reform era. The second section seeks to explain the pattern and impact of payment reforms in contemporary China.

Together, what do these theoretical explorations tell us about the incentive structure of China's healthcare delivery system? Many of the standard results of health economics of course apply: FFS gives incentive to over-supply the most profitable services; high supply-side cost sharing, such as fully prospective payment, gives incentives to reduce spending in both efficient and inefficient ways (stinting, risk selection, and so on). Administered prices provide distorted signals to providers, not reflecting true scarcity of resources and marginal social value, as market-based prices in a transparent and well-functioning market might. Providers are far from immune to financial incentives, and will respond in predictable ways.

Applying these concepts to China may have merit, but is not particularly novel. Digging deeper,

one might ask: why were incentives designed they way they were? Why does China have a problem of “*yi yao yang yi*” (supporting healthcare providers through drug sales)? Why did policymakers choose and legitimize drug sales as the profit center for providers throughout East Asia, but not necessarily in other developing or transitional economies (e.g. Eastern Europe) or elsewhere? To answer these questions, a comparative historical institutional analysis, such as the theory presented in the first part of the paper on physician dispensing and separation reforms, has particular value.

Patterns of incentives are not random. The incentive structure of China’s current healthcare delivery system has deep historical roots and institutional logic. The choice to make the profit margin on dispensing drugs positive seemed logical and appropriate to PRC policymakers in the 1950s because it reflected an institutional element that had long defined clinical reality. Even after a transformative revolution, important institutional elements from the past – especially shared beliefs about the proper form of a doctor-patient encounter and legitimate source of revenue for providers – shaped the options available. In the words of Greif (2006), there is a “fundamental asymmetry” between the structure inherited from the past and the technologically feasible alternatives. Specific conditions make a particular institutional arrangement self-enforcing or self-reinforcing. Recognizing these influences from the past, with their inherited logics, is not to say that the current incentive structure is optimal or unchangeable; far from it. Shifts of ‘quasi-parameters’ can render longstanding practices self-undermining (Greif 2006).

The large increase in available medical technologies in China over the last several decades, combined with the current expansion of insurance in China, both push toward separation of prescribing and dispensing (SPD). Introducing insurance lowers the costs of separation eventually, because provider payment need no longer be defined by patient willingness to pay linked to expectation of obtaining a drug during a doctor’s visit; patients are more willing to accept new ‘rules of the game’ when the rulemakers are the ones footing the bill. Insurance not only makes SPD ‘easier’ in this sense, it also makes SPD increasingly necessary. Without a strong demand-side constraint, SID can threaten the financial sustainability of social health insurance. Similar concerns spurred separation reforms in Korea and Taiwan.

In sum, theory does not predict SPD always and everywhere represents a “better” arrangement. The relative magnitude of the costs and benefits—and, more importantly, the perceived net benefit—of either arrangement depends on social and economic circumstances. The appropriate institutional arrangement for the production of medical services and drug products depends on

the relative scarcity of various inputs and the beliefs and norms of the participants. In societies with long traditions of indigenous herbal medicine and physician dispensing, like China, it may take large changes such as the rise of the welfare state and third-party payers to side-step and remold these patient and provider beliefs and norms. Such a process does not indicate that China is “backward” relative to societies with longer experience with SPD. But it does provide greater urgency for provider payment reforms that introduce appropriate levels of supply-side cost sharing, and suggests the usefulness of comparing experiences with regions sharing similar historical and cultural legacies.

References

- [1] Abbott, A. 1991. The order of professionalization: An empirical analysis. *Work and Occupations* 18, (4): 355.
- [2] Abbott, A. 1988. *The system of professions: An essay on the division of expert labor*. University of Chicago Press.
- [3] Abbott, A. D., and A. Abbott. 2001. *Chaos of disciplines*. University Of Chicago Press.
- [4] Afendulis, Christopher C., and Daniel P. Kessler. 2007. Tradeoffs from integrating diagnosis and treatment in markets for health care. *American Economic Review* 97, (3) (06): 1013-20.
- [5] Aoki, Masahiko. 2004. Towards a comparative institutional analysis: Motivations and some tentative theorizing. In *The foundations of the new institutional economics.*, ed. Claude Menard, 477-495. Elgar Reference Collection. International Library of the New Institutional Economics, vol. 1; Cheltenham, U.K. and Northampton, Mass.;; Elgar.
- [6] ———. 2001. *Toward a comparative institutional analysis*. Comparative Institutional Analysis Series; Cambridge and London: MIT Press.
- [7] Arrow, K. J. 1963. Uncertainty and the economics of medical care. *American Economic Review* 53, (5): 941–973.
- [8] Baker, G., R. Gibbons, and K. J. Murphy. 2002. Relational contracts and the theory of the firm. *Quarterly Journal of Economics* 117, (1): 39-84.

- [9] Chou, Y. J., W. C. Yip, C. H. Lee, N. Huang, Y. P. Sun, and H. J. Chang. 2003. Impact of separating drug prescribing and dispensing on provider behaviour: Taiwan's experience. *Health Policy and Planning* 18, (3) (Sep): 316-29.
- [10] Cochran, Sherman. 2006. *Chinese medicine men: Consumer culture in China and Southeast Asia*. Cambridge and London: Harvard University Press.
- [11] Cowen, D. L., and W. H. Helfand. 1990. *Pharmacy: An illustrated history*. Harry N. Abrams.
- [12] Du, L., 2005, *The Marketing of Traditional Medicines in China: The Case of Guangxi Province*. Lewiston, NY: Edwin Mellon Press.
- [13] Duggan, M.G., 2000, Hospital ownership and public medical spending, *Quarterly Journal of Economics* 115, 1343-1373.
- [14] Eggleston, K., 2008a, Soft budget constraints and the property rights theory of ownership, *Economics Letters* 100, 425-427. <http://dx.doi.org/10.1016/j.econlet.2008.03.010>.
- [15] Eggleston, K., 2008b, Comparative institutional analysis of the integration of diagnosis and treatment in health care markets: Evidence from physician dispensing in East Asia. Working paper, Asia Health Policy Program, Shorenstein Asia-Pacific Research Center, Stanford University.
- [16] Eggleston, K., 2008c. 'Kan Bing Nan, Kan Bing Gui': Challenges for China's Healthcare System Thirty Years into Reform, Chapter forthcoming in Stanford China Program book on contemporary China.
- [17] Eggleston, K., 2009. Editor. *Prescribing Cultures and Pharmaceutical Policy in the Asia-Pacific* (Shorenstein Asia-Pacific Research Center, Stanford University, through Brookings Institution Press: forthcoming in 2009).
- [18] Eggleston, K., and B.M. Yang, 2008, The Institution of Physician Dispensing in Comparative Historical Perspective: Theory and Evidence from Korea, forthcoming in *Prescribing Cultures and Pharmaceutical Policy in the Asia Pacific* (K. Eggleston, ed., Shorenstein Asia-Pacific Research Center and Brookings Institution Press, 2009).

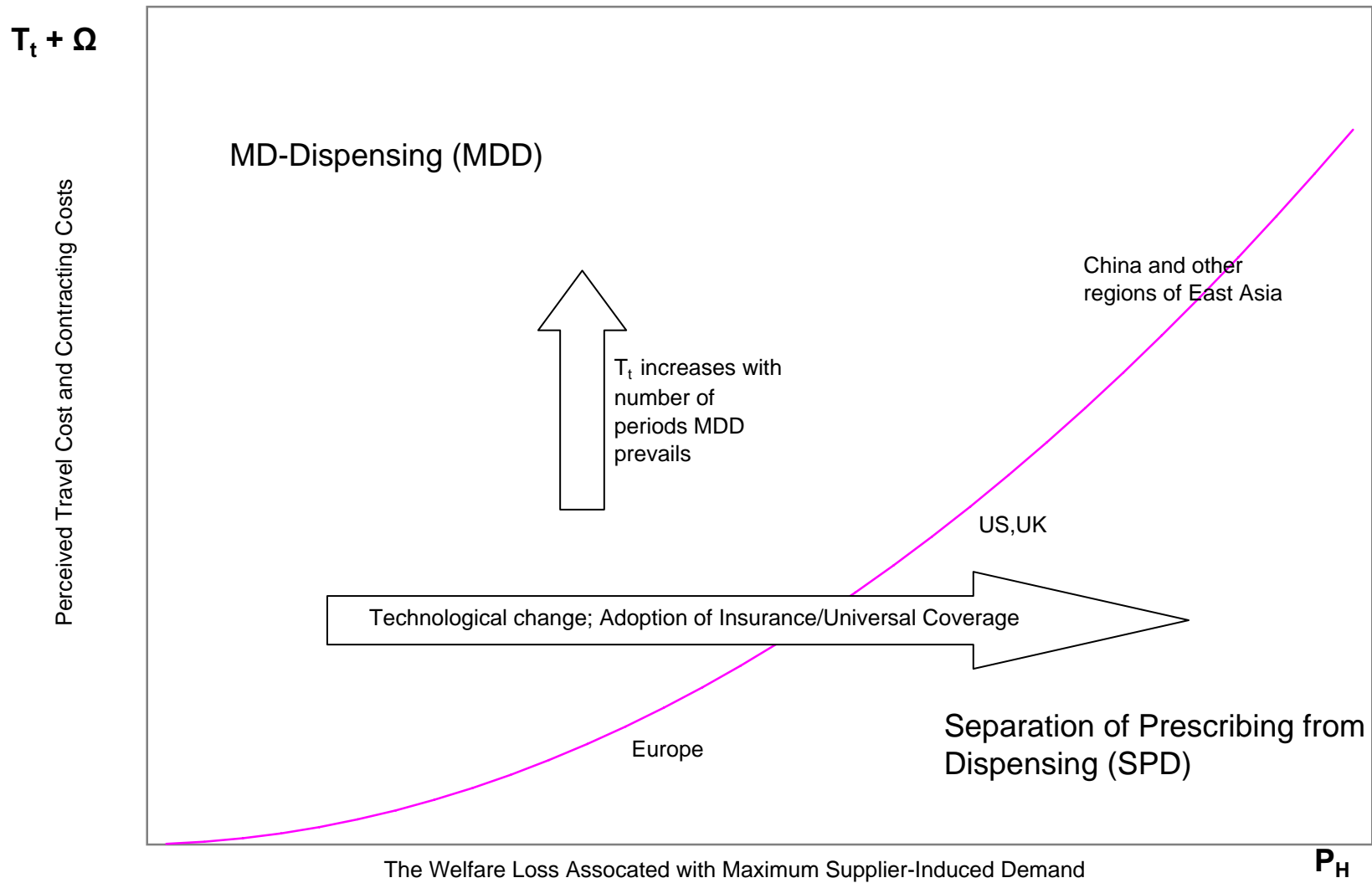
- [19] Eggleston, K., Y. Shen, M. Lu, C. Li, J. Wang, 2008a. “Soft budget constraints in China: Evidence from the Guangdong hospital industry.” Working paper, Shorenstein Asia-Pacific Research Center, Stanford University, September 2008.
- [20] Eggleston, K., M. Lu, C. Li, J. Wang, 2008b, “Physician Payment Incentives, Cost and Quality of Hospital Services: Evidence from Guangdong, China,” forthcoming in *China Labor Economics*, published by Institute of Population and Labor Economics, Chinese Academy of Social Sciences.
- [21] Fong, Yuk-fai. 2005. When do experts cheat and whom do they target? *Rand Journal of Economics* 36, (1) (Spring): 113-30.
- [22] Gaudl, R. 2005. *Comparative health policy in the Asia-Pacific*. Open University Press.
- [23] Greif, Avner. 2006. *Institutions and the path to the modern economy: Lessons from medieval trade*. Political Economy of Institutions and Decisions series; Cambridge and New York;; Cambridge University Press.
- [24] Gruber, Jonathan, and Maria Owings. 1996. Physician financial incentives and cesarean section delivery. *Rand Journal of Economics* 27, (1) (Spring): 99-123.
- [25] Hart, O. 1995. *Firms, contracts, and financial structure*. Oxford University Press.
- [26] Hart, O., and J. Moore. 1990. Property rights and the nature of the firm. *Journal of Political Economy* 98, (6): 1119-58.
- [27] Hart, O., S. Shleifer and R.W. Vishny, 1997, The proper scope of government: Theory and an application to prisons, *Quarterly Journal of Economics* 112, 1127-1161.
- [28] Helman, C. G. 2000. *Culture, health and illness*. Butterworth-Heinemann Boston.
- [29] Huang, Y., and Y. Yang, 2008. “Pharmaceutical pricing in China,” forthcoming in *Prescribing Cultures and Pharmaceutical Policy in the Asia Pacific* (K. Eggleston, ed., Shorenstein Asia-Pacific Research Center and Brookings Institution Press, 2009).
- [30] Hubbard, Thomas N. 2002. How do consumers motivate experts? Reputational incentives in an auto repair market. *Journal of Law and Economics* 45, (2, Part 1) (Oct.): 437-68.

- [31] ———. 1998. An empirical examination of moral hazard in the vehicle inspection market. *Rand Journal of Economics* 29, (2) (Summer): 406-26.
- [32] Iizuka, T. 2007. Experts' Agency Problems: Evidence from the Prescription Drug Market in Japan. *RAND Journal of Economics*, 38(3): 844-862.
- [33] Iizuka, T. 2008. "The economics of pharmaceutical pricing and physician prescribing in Japan," forthcoming in *Prescribing Cultures and Pharmaceutical Policy in the Asia Pacific* (K. Eggleston, ed., Shorenstein Asia-Pacific Research Center and Brookings Institution Press, 2009)..
- [34] Kim, Hak-Ju, and Jennifer Prah Ruger. 2008. Pharmaceutical reform in South Korea and the lessons it provides. *Health Affairs* July/August 2008; 27(4): w260-w269.
- [35] Kleinman, A. 1981. *Patients and healers in the context of culture: An exploration of the borderland between anthropology, medicine and psychiatry*. University of California Press.
- [36] Kornai, J., 1986, The softness of budget constraint, *Kyklos* 39, 3-30.
- [37] Kornai J, Maskin ES, and Roland G. 2003. Understanding the soft budget constraint. *Journal of Economic Literature* 41, 1095-1136.
- [38] Kwon, Soonman, and Michael R. Reich. 2005. The changing process and politics of health policy in Korea. *Journal of Health Politics, Policy and Law* 30, (6) (12): 1003-25.
- [39] Liu, X., and A. Mills, 2003. The Influence of Bonus Payments to Doctors on Hospital Revenue: Results of a Quasi-experimental Study, *Applied Health Economics and Health Policy* 2(2): 91-98.
- [40] Lundin, D. 2000. Moral hazard in physician prescription behavior. *Journal of Health Economics* 19, (5): 639-62.
- [41] Meng, Q. 2008. Case-based provider payment systems in China. Working paper submitted to "Provider Payment Incentives in the Asia-Pacific" conference. Shandong University Center for Health Management and Policy working paper.
- [42] North, D. C. 2005. *Understanding the process of economic change*. Princeton University Press.
- [43] ———. 1991. Institutions. *Journal of Economic Perspectives* 5, (1): 97-112.

- [44] ———. 1990. *Institutions, institutional change and economic performance*. Cambridge University Press.
- [45] Porter, R. 1997. *The greatest benefit to mankind: A medical history of humanity from antiquity to the present*. HarperCollins.
- [46] Rodwin, M. A., and A. Z. Okamoto. 2000. Physicians' conflicts of interest in Japan and the United States: Lessons for the United States. *Journal of Health Politics, Policy and Law* 25, (2): 343-76.
- [47] Santoro, M.A., and C.M. Liu, 2008. "The global threat to safety posed by China's underdeveloped drug regulation regime," in *Prescribing Cultures and Pharmaceutical Policy in the Asia Pacific* (K. Eggleston, ed., Shorenstein Asia-Pacific Research Center and Brookings Institution Press, 2009).
- [48] Shen, Y., and K. Eggleston, 2008, The effect of soft budget constraints on access and quality in hospital care. NBER Working paper #14256.
- [49] Sloan, F.A., 2000, Not-for-profit ownership and hospital behavior, in: A.J. Culyer and J.P. Newhouse, eds., *Handbook of Health Economics*, Vol. 1B (North-Holland, Amsterdam) 1141-1174.
- [50] Sonnedecker, G. 1986. *Kremers and Urdang's history of pharmacy*. American Institute of the History of Pharmacy.
- [51] Starr, P. 1982. *The social transformation of American medicine*. Basic Books.
- [52] Sun, Q., M.A. Santoro, Q. Meng, C. Liu, and K. Eggleston, "Pharmaceutical Policy in China," *Health Affairs* 27, no. 4 (2008): 1042–1050; 10.1377/hlthaff.27.4.1042.
- [53] Taylor, Curtis R. 1995. The economics of breakdowns, checkups, and cures. *Journal of Political Economy* 103, (1) (Feb.): 53-74.
- [54] Tomita, N. 2008. "The political economy of incrementally separating prescribing from dispensing in Japan," in *Prescribing Cultures and Pharmaceutical Policy in the Asia Pacific* (K. Eggleston, ed., Shorenstein Asia-Pacific Research Center and Brookings Institution Press, 2009).

- [55] Unschuld, P.U. 1986. *Medicine in China*.
- [56] Unschuld, P.U. 1987. Traditional Chinese medicine: Some historical and epistemological reflections. *Social Science & Medicine*, 24, (12): 1023-9.
- [57] Wolinsky, Asher. 1993. Competition in a market for informed experts' services. *Rand Journal of Economics* 24, (3) (Autumn): 380-98.
- [58] Worling, Peter. 2005. "Pharmacy in the early modern world, 1617 to 1841 AD," chapter 5 in *Making Medicines: A Brief History of Pharmacy and Pharmaceuticals*. Stuart Anderson, ed. London: Pharmaceutical Press, 2005.
- [59] Zhang, Jiale, 2007, A DID analysis of the impact of health insurance reform in the city of Hangzhou, *Health Economics* 16(12), 1389-1402.
- [60] Zhang, Jiale, 2007, The impact of a diagnosis-related group-based prospective payment experiment: The experience of Shanghai, unpublished manuscript.

Figure 1.
MD-Dispensing (MDD) and Separation of Prescribing from Dispensing (SPD)
as Self-Enforcing Institutions



Total exp. prescription, % total expenditure on health, TEH

OECD Health Data 2008

