



**NEW ZEALAND INSTITUTE FOR THE STUDY
OF COMPETITION AND REGULATION INC.**

Financial Risk in Primary Health Care Contracting: Implications for Sector Structure, Ownership and Outcomes

May 2007

Bronwyn Howell

New Zealand Institute for the Study of Competition and Regulation Inc. and
Victoria Management School, Victoria University of Wellington, PO Box 600,
Wellington, New Zealand. Email bronwyn.howell@vuw.ac.nz

***Acknowledgement:** The author wishes to acknowledge the helpful comments provided by Lewis Evans, the support of Glenn Boyle and the New Zealand Institute for the Study of Competition and Regulation, whilst undertaking research into New Zealand primary health care markets and the Deane Endowment Trust for financial support during the preparation of this paper. Any errors or omissions remain the responsibility of the author.*

Abstract

The success of supply-side risk-sharing contracts in achieving behavioural change amongst health care providers is dependant upon the trade-off between reduced costs from reduction in ‘predictable risks’ (e.g. supplier-induced demand) and increased costs from the sharing of ‘random risk’, optimally managed under a single large pool into several smaller risk-bearing pools. Typical capitation contracts do not distinguish between the degree of random and predictable risk shared with service providers. The strength of the financial incentives in capitation contracts must be finely balanced, to ensure that elements associated with ‘luck’ do not lead to perverse outcomes that crowd out the achievement of the desired benefits.

By examining the effects of financial risk-sharing on incentives for practitioner-owner behaviour and practice ownership, this paper explores some of the likely consequences of contracts sharing over-large amounts of random risk with primary health care practitioners. Likely perverse outcomes include fewer and higher-cost consultations than under fee-for-service, a systemic allocation of care quality penalising sicker-than-average populations that occurs in addition to the well-established effects of deliberate cream-skimming, a bifurcation of the supply and allocation of services into a two-tier system of small privately-owned providers serving healthier-than-average populations and large nonprofit providers serving sicker-than-average populations, and long-term distortions in the allocation of practitioners amongst practice ownership forms and systems of different design. The likelihood of these outcomes occurring and the costs they invoke must be carefully considered by policy-makers when designing remuneration in primary health care contracts.

Introduction

“There are many mechanisms for paying physicians; some are good and some are bad. The three worst are fee-for-service, capitation and salary”. (Robinson, 2001:149)

Recent years have seen the emergence of many health care policies employing risk-sharing tools aimed at aligning the activities of service providers with the objectives of purchasers and funders. An example is managed care, where responsibility for managing all aspects of the financing and delivery of health care for a defined set of patients is assigned to specific care management entities (Robinson, 2004; Rivers and Tsai, 2003). Whilst managed care models involve the use of many non-price elements (e.g. utilisation review), a common characteristic is some element of financial risk-sharing (‘supply-side cost-sharing’ – Ellis and McGuire, 1986).

The agency theory rationale for supply-side cost-sharing is that by contractually linking the care manager’s financial returns to a proxy for the achievement of some desired characteristics, or decoupling of financial returns from undesirable characteristics, the desired activities will be pursued and the undesirable ones eschewed. Effort exerted by the practitioner, which is unobserved and unobservable by the policy-maker or insurance purchaser setting the contract terms, will be redirected towards the pursuit of desirable, rather than undesirable, activities (Robinson, 2001). Commonly-used supply side cost-sharing contracts are price-and-volume, partial and full capitation (Danzon, 1997).

Partial or full capitation contracts have become common in primary health care remuneration. They comprise the principal form of government-funded remuneration for England’s Primary Care Trusts (Keen, Light and Mays, 1999), are very common ways for insurers to remunerate primary care physicians in the United States (Robinson, 2004; Hagen, 1999) and form the basis of New Zealand’s partially tax-funded primary health care strategy implemented in 2002 (King, 2001). Decoupling remuneration and inputs (e.g. physician time, consultations) and linking remuneration with outcome proxies (e.g. patients, who if healthier generate fewer costs for the care provider as they consume fewer consultations) has led to their extensive use in public sector ‘outcomes-based’ (Honore, et. al, 2004) and ‘performance-based’ (Martin, 2002) contracting generally, as a method for health care purchasing. Reduced emphasis on the consultation as the payment determinant is attributed with a shift in primary care delivery focus away from interventions in the event of illness towards the promotion and maintenance of wellness (Coster and Gribben, 1999; Cumming, 1999; Malcolm, 1997). Crampton, Sutton and Foley (2001) summarise the justification for New Zealand’s primary care policy adoption

of capitation to be increased equity, targeting of high health need, encouraging a team approach to primary health care, and a change in focus towards a care management model as opposed to a model of episodic intervention with a focus on illness.

Many commentaries and analyses of health care capitation contracts take a broad view over many dimensions of how incentive contracts affect provider behaviour (e.g. Robinson, 2001; Dranove, Simon and White, 1998) and the relationship between insurer-purchasers and service providers (e.g. Cooper and Rebitzer, 2006). Others have focused principally on the effects for government and patients (e.g. Crampton, Sutton and Foley, 2002; Cumming, 1999; Coster and Gribben, 1999; Malcolm, 1997). Few focus upon how the contracts affect the incentives for ownership of primary health care firms and thereby influence sector structure. Dranove, Simon and White (2002) provide one exception, albeit addressing primary care amongst their analysis of the effects on specialists and hospitals. This gap in the literature is surprising, given the extent to which in the primary care sector the individuals whose behaviour is to be changed are generally the owners of the businesses as well as service providers (Scott, 2000; Dranove and Satterthwaite, 2000).

This paper addresses the gap in the literature by focusing specifically on the role of financial risk in primary health care contracts, its effects upon sector ownership and investment, and how these effects flow through into the supply of primary care services, the distribution of available services amongst patients and overall sector structure. At its core is the fact that all contracts allocate financial risk between the purchaser and the service provider. The effectiveness of any risk-sharing agreement relies upon how well the rewarded proxy matches the desired objective. If the match is close, then the desired outcome is highly likely to occur. However, if the match is poor, or fails to anticipate possible provider responses, the contract opens up the potential for perverse outcomes to arise, the costs of which may exceed the gains anticipated from pursuit of the desired behaviour (Dranove, Simon and White, 1998; Robinson, 2001).

Specifically, the paper examines the ways in which various contract forms manage the financial consequences of ‘uncontrollable’ and ‘unpredictable’ factors in addition to the ‘controllable’ factors of provider behaviour that the risk-sharing contract seeks to address. A feature of health care contracting is that third-party purchasing occurs because governments and insurers are more efficient managers of the financial risks of patients’ unknown demand for health care than the patients themselves. When third-party purchasers enter into capitation contracts with service providers, they share random patient demand risk along with the costs of predictable risk that are controllable by the providers. These risks are not shared under fee-

for-service or price-and-volume contracts. As the capitation contract bundles both risks together and shares them in equal proportions, the ability for capitation contracts to successfully alter provider behaviour without inducing over-costly effects arising from reduced efficiency in managing random risk lies in setting the incentive strength appropriately. If the incentive is too strong (i.e. shares 'too much' random risk), then there will inevitably be consequences that flow through into sector costs, ownership patterns and structure. A number of likely perverse outcomes in the event of over-much random risk being shared are identified.

The paper proceeds as follows. Section one discusses the contracting process, and the ways in which contracts that allocate risk affect sector efficiency, the firm survival prospects and the willingness of individuals to invest their physical and human capital. Section two applies the theory from section one to primary health care markets. This section discusses patient demand uncertainty, third party purchasing in health care, and how risks of different types affect the behaviour of primary health care providers differently. Section three then discusses some of the likely consequences arising when over-much random risk is shared with primary care providers. These include fewer and higher-cost consultations than under fee-for-service, a systemic allocation of care quality penalising sicker-than-average populations that occurs in addition to the well-established effects of deliberate cream-skimming, a bifurcation of the supply and allocation of services into a two-tier system of small privately-owned providers serving healthier-than-average populations and large nonprofit providers serving sicker-than-average populations, and long-term distortions in the allocation of practitioners amongst practice ownership forms and systems of different design. Section four concludes with a discussion of the implications for policy-makers and insurance companies.

1. *Contracts and Risk*

A fundamental economic axiom is that a firm will be financially viable in the long run only if the income received from the provision of goods and services (or gifts, donations and sponsorship to fund the provision of benefits in the case of a charitable organisation) at least equals the costs of producing those goods and services. Costs include a fair return to the owners for the use of the physical and human capital provided, or the costs of replacing essential capital (both physical and human) in the case of charitable and non-profit organisations with no defined shareholder-owners. If income does not exceed these costs, in the long run the firm must either increase the prices it charges, reduce its costs or exit from producing goods and services (including charitable ones). The owners of human and physical capital will receive better returns on their capital if it is invested elsewhere.

For a price-taking firm operating in a competitive market, or operating in a market where price regulations or concentration of buyer power inhibit its ability to raise prices unilaterally, increasing the income received per unit of output is not feasible. The firm's survival depends upon achieving cost reductions. Cost reductions may occur from innovations in the firm's production processes that lead to lower unit costs per quantity of output of a given quality produced, a reduction in the number of goods and services produced, or a reduction in quality of goods produced.

Self-evident though they may be, these simple concepts are fundamental to the structure of an industry sector – who will own the firms (i.e. supply the capital) and how they will interact with other parties. Contracts affect financial outcomes for firm owners, and thus their willingness to invest in particular firms, the activities that the firms will engage in, and the types of contracts that characterise those activities. Different contracts will therefore lead to different ownership incentives and firm behaviour, in a complex nexus of sector interrelationships, just as a change in owner will affect the nature of the contracts a firm will enter into. Irrespective of whether a firm is maximising financial profit or pursuing some other objective, the bottom line is that the primary objective of the owners and managers of any firm is to first ensure that it survives, if it is unviable to wind it up as rapidly as possible to avoid incurring further unrecoverable losses, and to ensure that, if the firm is viable, then the activities it undertakes do not increase the risks to the firm's survivability beyond its capacity to bear them. Parties with market power must be aware of the consequences that their contracts will have on firm survivability, investment patterns and industry structure. The consequences of injudicious contracts may ultimately impact upon the ability for the party with market power to achieve its desired objectives.

1.1 The Contracting Process

When a firm is taking all possible steps to utilise internal productive efficiency enhancements, its survivability is determined by the contracts it enters into with customers and suppliers. In negotiating the contract terms, managers will endeavour to ensure that the firm's long-term survival prospects are enhanced (e.g. by only entering into sale contracts where remuneration at least meets costs). A typical contract process will involve search for potential partners, negotiation of terms, drawing up of the contract, monitoring contract performance and enforcing the terms of the contract in the event of breach (Milgrom and Roberts, 1992). Parties will voluntarily enter into a contract only where its terms result in both parties being better off than if the contract did not exist (Coase, 1937).

In order to ensure that the contract terms are advantageous to both parties, the parties must both assess the specification of the goods and the price of exchange, and anticipate additional costs and risks arising as a consequence of entering into the agreement. Parties must agree on terms that minimise transaction costs, use incentive terms and monitoring to limit opportunism, allocate risk, facilitate investment in specific assets and allocate property rights in such a manner that both parties are better off with the contract than without it (Boyd, Evans and Quigley, 2000, building on Williamson, 1986).

1.2 Uncertainty, Risk Allocation and Contract Efficiency

Establishing the terms dealing with the management of transaction costs, facilitation of investment in specific assets and allocation of property rights is usually straightforward, as these relate principally to deterministic factors (although one party may have stronger bargaining power, which may affect allocation of gains). However, management of opportunism and allocation of risk are more problematic, as they deal with uncertainties. For example, will a party act opportunistically by using superior information to deprive the other party, and if so, how will this occur?; Will the ‘state of nature’ change, altering the assumptions under which the contract was entered into, and thereby changing the payoffs received by both parties?). Contractual uncertainty thus creates a risk to the long-term viability of the firm. Two types of risk must be considered: unpredictable and uncontrollable changes in the state of nature that neither party can anticipate or manage (‘random’ risk), and factors that may arise in the future, but are predictable and controllable by one or other of the parties to the contract (‘predictable’ risk) (Milgrom and Roberts, 2002).

An efficient contract will allocate the costs of risks eventuating to the party best able to bear them. Contract theory indicates that desired outcomes are achieved at least cost when (Milgrom and Roberts, 1992):

1. the costs of ‘predictable’ risk are borne by the party who can best control the level of the risk occurring, via standard incentive terms (e.g. the costs of shirking should be borne by the party who can choose to either shirk or exert full effort); and
2. ‘random’ risks are allocated in proportion to the ability of each party to bear them (e.g. a well-resourced party can ‘insure’ a less well-resourced party against financial failure in the event of random events by absorbing the costs of random risk, albeit with the premium for such insurance reflected in the contract compensation).

In the event of either risk type being inappropriately allocated, the contract outcome is less advantageous than it might otherwise have been for at least one of the parties. For example, if a less well-resourced firm incurs costs of a ‘random’ risk eventuating that exceed its revenues

from contract remuneration and reserves, then at best it must reduce costs by reducing service or product quality, or at worst it will fail financially and exit the market. In either event, the more well-resourced party also suffers losses (e.g. under the contract, it receives lower-quality products than specified, or no products at all) and must incur additional costs to seek redress (monitoring and enforcement) and the search and contracting costs relating to seeking and negotiating a new contract with a new supplier.

If the costs to the well-resourced party of seeking redress are higher than the costs of bearing the losses on the contract, then it would have been less costly to have underwritten the other party's random risk via the contract, and continue to enjoy high quality products and services should unforeseen events occur (with the costs of bearing this risk reflected in the contract compensation terms). Likewise, premiums paid for high quality (or sanctions for low quality) make it costly for the producing party to consciously choose to lower quality, thereby disadvantaging the purchasing party. It is more efficient for the purchasing party to manage the risk of quality reduction by offering incentives for high quality, up to the level of the costs anticipated in seeking redress for low quality.

2. *Application to Primary Health Care*

The principal contract in the provision of primary health care is between the purchaser of care and the service provider. The nature of the health care product, and the costs and risks associated with its exchange, have led to a sector structure characterised by large third-party purchasers contracting with small, mostly owner-operated practices, to deliver health care to patients.

2.1 *The Service Provider*

Primary care service providers may be large corporate entities, but are more commonly sole general practitioners or a small group of practitioners working co-operatively (often with a formal co-operative agreement, in the form of Hansmann's (1996) supplier-owned co-operative) in order to deliver services to a defined set of patients with whom they have either explicit contracts or implicit agreements to provide ongoing care delivery as indicated by the patient's health state (the 'patient list', 'book' or 'patient panel'). The repeated interactions between a specific individual and primary care provider also distinguish this sector from other forms of health care delivery (e.g. surgical intervention) where repeat transactions by a single individual with a specific practitioner are much rarer (Scott, 2000; McGuire, 2000).

Both Robinson (2001) and Scott (2000) identify the ‘one-to-one’ nature of general practice, and Newhouse (1973) identifies the advantages financially of sole practice in primary care. Sole practice is generally preferable as minimal teamwork is required to deliver any specific instance of care (compared to surgery, where many different specialists work concurrently on the same patient) (Newhouse, 1973), there are limited capital investments (aside from human) required, leading to limited need to form corporate entities for the purpose of capital-raising (Howell, 2005), and because the practitioner is able to receive the rewards from exerting his own effort and returns on his own human and physical capital without having to share them with a partner whose effort is unobservable and therefore difficult to contractually assign under a partnership or equity-sharing agreement (Hart and Grossman, 1986). Dranove and Satterthwaite (2000) identify the preference of both patients and practitioners for direct contractual relationships between patients and specific practitioners, as opposed to a firm, for the ongoing delivery of care. Under these arrangements, the individual practitioner ‘owns’ the patient list and is personally incentivised to exert effort to build relationships with patients and to repeatedly deliver high-quality care in order to secure future income streams (crystallised as practice ‘goodwill’ on the balance sheet). As such effort is extremely difficult to incentivise contractually, all other things being equal, individual ownership is economically superior to salary or partnership, as the practice owner directly receives the return to his own effort exerted via returns as owner. The typical practice is thus small (between 1200 and 2000 patients is common in New Zealand – Howell, 2005), with consultations being delivered repeatedly to the same small number of individuals.

The superiority of the sole practitioner and provider co-operative arrangements appears confirmed by their persistence over time (similar ownership arrangements prevail in other markets where a high proportion of capital invested is human, such as law and accountancy). If any other model of ownership was economically or clinically superior, then practitioners would have voluntarily adopted it or investor-owners who were not themselves practitioners would have entered the market, hiring salaried employees in order to appropriate the gains from a more efficient ownership form (Hansmann, 1996). Dranove, Simon and White (2002) identify that the rise of corporate primary health care provision is strongly correlated with the rise of capitation contracting. This paper will demonstrate that this is likely occurring as a consequence of the ways in which capitation contracts allocate financial risk to small firms with little physical capital or debt-bearing capacity, and because superior profits are available in some practices. However, the mechanism via which this occurs is not the pursuit of larger profits, but the amelioration of losses.

2.2 Provider Cost Structures

The typical product delivered by primary care providers is a patient consultation. It is also the provider's primary cost-driver. Irrespective of whether the motivation for the consultation is treatment- or prevention-related, in a typical service the patient has an exclusive meeting (e.g. face-to-face, telephone) with the practitioner or other staff member (e.g. nurse). All consultations require inputs of time and knowledge (human capital) by the practitioner and other staff.

Consultation costs are principally comprised of the opportunity cost of the practitioner's and other staff member's time (e.g. nurse, receptionist) (including a fair return on human capital for the practitioner-owner). As physical capital requirements for a primary health care business are generally low, the share of overheads in average consultation costs is small, comprising principally of rent, power, supplies, etc. Cost reductions per consultation are generally achievable only by reducing the time per consultation or the practitioner's expectation of the acceptable return received on (predominantly) human capital (i.e. reservation wage).

2.3 Random Risk Management and Third-Party Purchasing

The purchaser of primary health care may be an individual, but is more commonly an insurance company or government agency purchasing care on behalf of a policy-holder or citizen. The industry is thus characterised by two distinct sub-markets: an upstream risk management (insurance) market, managing individuals' uncertainties regarding the need for and ability to finance care, and a downstream health care provision (service delivery) market where the care for ill individuals¹ is purchased and delivered (Howell, 2006).

The upstream risk management market arises as a consequence of the uncertainty that all individuals face with respect to their ability to finance care when falling ill. For the most part, knowing if and when a given individual will fall ill is unpredictable (i.e. 'random'). The ways in which health sector contracts handle this uncertainty (hereinafter termed 'random demand risk') are critical to sector efficiency. United States evidence suggests that only around 20% to 25% of individual demand variation can be reliably predicted using characteristics such as age, ethnicity, income, gender and previous consumption of care (Newhouse, 1996; Robinson, 2004).

¹ For the purposes of this paper, 'falling ill' means a patient requires a consultation. The consultation may be an intervention for a specific illness, or the requirement that some form of preventative treatment (e.g. vaccination, education) be delivered.

However, for a large number of people, the probability of a predictable number falling ill in a given time period is ascertainable, even though the actual identity of the individuals who will fall ill is unpredictable. This leads to demand for insurance. Arrow (1963) identifies that individuals can enter into contracts that aggregate (pool) individual uncertainties about future demand for care in order to manage ‘random demand risk’ more efficiently than if the individual managed the risk alone (self-insurance). The pool may be managed by a firm or a government. In either case, the pool ‘insures’ the individual against unpredictable costs. Via explicit or implicit² insurance contracts, premiums paid when well or taxation funding agreements guarantee certainty of access to care (as specified in the insurance contract or government policy terms) for an individual when illness strikes (Zeckhauser, 1970; Cutler and Zeckhauser, 2000). The financial vulnerability of an individual to the costs of random events (risk-aversion) motivates individuals with low ability to bear the random costs to share the risk by buying insurance policies or by supporting taxation-funded government schemes underwriting health care costs.

The superiority of pooled management (insurance) over individual management (self-insurance) derives from two sources. Firstly, the law of large numbers enables better ex ante predictability of the likely demand and hence costs of care for a group of individuals than for the individual alone. The insurer can allocate the costs of care for those individuals falling ill amongst all pool members, removing individual uncertainty about affordability should the ‘bad luck’ of falling ill occur. Secondly, the larger the size of the group, the more individuals amongst whom the costs of differences between the manager’s predictions and actual costs of care incurred can be shared, thereby reducing the cost per individual of ‘random’ variations arising from the limited foresight of the insurer in predicting actual costs. The pool manager can also use surpluses (reserves) from ‘lucky’ periods where demand (and cost) was lower than anticipated to cover losses from ‘unlucky’ periods when demand (and cost) exceeds predictions. The ability to levy members in future periods provides security for loans to cover losses in periods where reserves are insufficient. Large insurers and governments are therefore in a much better position to manage the costs of random demand variations more cost-effectively than risk-averse individuals. As most risk-averse individuals prefer to sacrifice a small, certain amount regularly to offset the financial consequences of an unpredictable, large, potentially financially crippling expense, they assign risk management, and purchasing responsibility to these third parties.

² Taxation-funded government pools undertake the insurance role implicitly as a consequence of policies and political decisions. The terms of the implicit contract are the obligations on both parties under the policy or legislation. Private sector insurance companies have explicit contracts with individuals identifying the obligations on both parties (the ‘insurance policy’).

2.4 Risk and Remuneration in Primary Care

Remuneration contracts for primary health care provision have typically been one of two types: simple fee-for-service and price-and-volume, where primary care providers are paid a fixed fee per patient consultation; and more complex contracts, such as partial and full capitation, where providers face incentives to manage the costs of care delivery because part or all of their remuneration is invariant to the number of consultations provided (Dranove and Satterthwaite, 2000; Danzon, 1997; Newhouse, 1996).

2.4.1 Risk Under Fee-for-Service

Under fee-for-service payment arrangements, primary care providers receive a fixed fee for each consultation provided (with different fees for different consultation types), reflecting the importance of the consultation as the principal cost driver. The consultation is an easily-measured metric, involving relatively low transaction and monitoring costs for both the purchaser and the provider. As long as the service fee at least meets the costs of providing the consultation, service providers will be financially viable in the long run. The provider faces no financial risks from the occurrence of truly random events affecting the number of consultations provided, such as the emergence of an unpredictable or uncontrollable epidemic (e.g. ‘bird flu’) or ‘random demand risk’ allocation (i.e. the natural distribution of patients amongst practices that results in some practices having a patient mix of individuals with unknown higher-than-average demand (‘sicker’) and some with a patient mix of lower-than-average demand (‘healthier’) individuals). Each consultation provided is fully compensated³. Random risks are borne by the purchasing entities. If more consultations are provided as a consequence of random events increasing practice demand, the additional fees are paid by the purchaser. It is the purchaser’s financial viability that is threatened by random risk variations.

The number of consultations provided by a practice remunerated by fee-for-service is determined principally by the willingness to supply practitioner and staff time at the contract fee. As the costs of each consultation are fully compensated, practices can supply as many or as few consultations as they chose (or register ‘on the books’ the number of patients they deem will result in the number of consultations the practice is willing to provide) without altering the ongoing financial viability of the practice. Incremental income from an additional consultation is at least equal to incremental cost, meaning financial viability (profitability) is unaffected by the level of effort expended (number of consultations provided).

³ Each consultation invokes an obligation to pay. Whilst all practices have bad debts – that is, patients who do not pay their bills - this is no different to any other business. It is at the discretion of the owner whether these bad debts are written off or pursued.

Whilst insulating providers from random risk, fee-for-service contracts invoke a specific predictable risk: supplier-induced demand. As both patients and their funding agents face an information asymmetry regarding the nature of and the most cost-effective treatment methods for treating, an illness, providers can request more or more costly tests and pharmaceuticals and/or induce the patient to consume more consultations than is necessary to effect a cure (Arrow, 1963; Pauly, 1968; Zeckhauser, 1970; Zweifel and Manning, 2000).

If the payer is an individual patient, supplier-induced demand is unlikely to be a significant problem, as a patient with limited reserves will quickly reach a point where the additional cost threatens his financial viability, and he will terminate the contract. However, if the payer is an insurer or government entity, the individual patient does not directly bear the costs of supplier-induced demand. The insuring entity pays the service provider. The patient faces little financial incentive to constrain the doctor's high cost-causing behaviour. Third-party purchasing also encourages the patient to engage in another well-known behaviour that increases predictable risk for purchasers – over-consumption of services (e.g. seeking subsidised treatment for very minor complaints that might be more cost-effectively treated by other means such as an over-the-counter pharmaceutical, or simply for reassurance – known as over-consumption by the 'worried well') (Pauly, 2000; Newhouse, 1996). In either case, third-party payers initially bear the additional costs of both of these forms of predictable risk, although all patients will ultimately collectively bear the costs via higher premiums or taxes.

The efficient contractual management of predictable risk involves making the party controlling the extent of the risk bear the costs of invoking it. If the practitioners inducing unnecessary consultations or ordering over-costly procedures, tests and pharmaceuticals are made to bear some of the additional costs of their excessive cost-causing activities, then the quantity of unnecessary consultations or over-costly tests and pharmaceuticals provided will reduce, thereby lowering the burgeoning costs levied ultimately upon policy-holders and taxpayers. Consequently, insurers and governments now routinely enter into supply-side cost-sharing agreements with practices in an endeavour to constrain demand and costs (Ellis and McGuire, 1986; Robinson, 2004; Ma and McGuire, 2002; Ma and Riordan, 2002; Cutler and Zeckhauser, 2000; Danzon, 1997; Newhouse, 1996 amongst others). Likewise, if patients are required to bear part of the costs of each consultation (via an excess or co-payment), then their propensity to engage in over-consumption may be muted, but not eliminated (Pauly, 1968; Pauly, 2000).

2.4.2 Predictable Risk Under Supply-Side Cost-Sharing

Capitation and price-and-volume contracts both directly address the predictable risk of too many consultations being provided under fee-for-service by imposing a cap on the quantity of services provided, either explicitly (price-and-volume) or implicitly (capitation). If practitioners choose to provide more or higher-quality services than anticipated in the contract, practice profits are reduced and long-term financial viability is threatened. However, practices keep any savings from cost-reducing innovations as profits. Thus, incentives for the pursuit of cost-reducing innovation are greater under supply-side risk-sharing than fee-for-service contracts. Capitation contracts have been used successfully in New Zealand to cap physician-controlled primary sector pharmaceutical and laboratory test spending (Malcolm, 1997).

Whilst addressing the fee-for-service over-provision risk, supply-side risk-sharing contracts invoke further predictable risks. Practices become more profitable under supply-side risk-sharing by reducing costs in a manner contrary to purchasers' intentions (e.g. providing lower-than-optimal service quality, such as shorter consultations, thereby reducing the benefits accrued by patients). These additional risks are typically managed by additional incentives for quality or more extensive monitoring to ensure that quality is not compromised. Although these additional requirements may increase explicit purchaser costs, they are justified if the cost of the risks avoided exceeds the additional costs of monitoring and quality incentives. In large part, the rise in managed care reflects increased endeavours by third-party payers to constrain the 'predictable' costs that the insurance risk management market creates (over-consumption from treatment subsidies).

2.4.3 Random Risk Under Capitation

The treatment of random risk under capitation is fundamentally different under capitation than under price-and-volume and fee-for-service. Capitation contracts pay a fixed fee f in respect of all individuals n for which the practice is responsible, irrespective of their actual demand for consultations. Under full capitation, the practice revenue is fixed at $F = fn$. Under partial capitation, a lower fixed fee is paid, and a variable fee v is paid ex post for each consultation. For q consultations delivered, practice revenue R is given by $R = F + vq$. When the cost per consultation is c , profit P is $P = F - q(c - v)$. Average Revenue per consultation is:

$$AR = R/q = (F + vq)/q \quad (1)$$

and Average Profit per consultation:

$$AP = P/q = F/q + (v - c) \quad (2).$$

Under fee-for-service and price-and-volume, the practitioner receives no fixed component ($f = 0$) determined ex-ante, but is remunerated ex post with an average revenue per consultation of v . Average profit per consultation is:

$$AP = (v - c) \quad (3).$$

The distinction between fee-for-service and price-and volume is that the number of consultations remunerated under equation (3) is restricted to an agreed number $q = Q$.

Equations (1) and (2) illustrate clearly that, whilst practitioner choices may affect the number of consultations provided under capitation, so too will random unknown and unpredictable influences upon the number of consultations demanded (random demand risk). By contrast, equation (3) illustrates that neither random nor predictable risk alter average revenues and average profits per consultation under fee-for-service and price-and-volume, as these are invariant to changes in q .

Financial viability of capitated practices is thus subject to factors uncontrollable by the practitioners. Two practices with patients of identical ex-ante characteristics funded identically by capitation and taking identical steps to manage predictable risk face different risks to financial viability simply because one was 'lucky' (i.e. its patients were 'healthier than average' as a consequence of unknown and unknowable factors not included in the 'predictable' factors determining contract compensation, or patients of this practice were not exposed to random demand shocks) and the other 'unlucky' (i.e. unpredictably 'sicker than average' patients, or exposed to random demand shocks). To the extent that fixed capitation payments f are adjusted for differences in ex-ante assessment of likely future demand differences between individuals, such payments are capturing only that extent of individual demand that is able to be converted to predictable risk via the algorithms employed by the insurer. The remainder is still unpredictable. If, as identified above, it is likely that over 70% of demand variation is unpredictable, all capitated practices will inevitably be subject to substantial random demand risk arising from the distribution of patients amongst practices, as well as any other random shocks to demand that may occur (e.g. an outbreak of a new virus, food poisoning, etc.).

When third-party payers enter into capitation contracts with service providers, they share both the 'predictable' contract risks arising from the presence of insurance and subsidised treatment, and some of the random demand risk which motivates patients to confer responsibility for managing this risk onto aggregated third-party risk-bearing entities in the first place. In effect, capitation contracts result in primary care providers becoming underwriters (insurers) of random demand risk. This confers on capitated providers all of the

responsibilities for managing reserves to meet variation in demand both in a single time period, and between time periods. Hagen (1999) indicates that regulating provider-insurers to ensure reserves or alternative risk management activities (e.g. purchase of reinsurance) is a very complex task, even in markets such as California with long experience of capitation contracts and sophisticated insurance regulation processes in place.

Under typical insurance arrangements, taxation or premium payments made by 'lucky' individuals can be used to pay the treatment costs of 'unlucky' ones. The larger the patient pool, the greater the probability that the number of 'lucky' individuals closely matches the number of 'unlucky' ones. The large pool has a greater likelihood of breaking even, or if profits or losses are made, they will be comparatively small. If the same large pool is fragmented into many smaller pools, then the probability of a close match between the number of 'lucky' and 'unlucky' individuals in each pool becomes much less. Some pools will be extremely lucky, making large profits, and others catastrophically loss-making. Unlike in the large pool, profits from 'lucky' small pools cannot be used to compensate the losses from 'unlucky' small pools. 'Unlucky' pool owners must use their own resources to meet the losses. 'Lucky' pool owners become beneficiaries as they are able to extract the profits granted to them by the random effects, and 'unlucky' pool owners must contribute additional resources equal in total to the profits extracted by the 'lucky providers' in order to stay financially viable. Thus, the cost of random risk management is substantially greater with a fragmented pool than with a single pool. American research suggests that capitated primary health care pools of fewer than 25,000 registered individuals face significant risks of financial failure (Hagen, 1999; Robinson, 2004).

If the gains and losses were truly random, then on average across time, gains incurred by a provider in some periods would cancel losses in others. However, the 'random' effects shared with practitioners also include the errors in cost predictions and any other unknown cost-causing events that have not been factored into the contract remuneration. They are 'unknown' only to the extent that the algorithms predicting likely demand for an individual ex ante are imperfect proxies for the individual's actual demand because they cannot accurately capture all relevant contributing influences. These unknown factors (e.g. unknown genetic predispositions) are likely to be highly correlated with the individuals in the patient pools served by each practitioner. As primary care patient pools are typically fairly stable between periods, loss-making and profit-making tend to become highly correlated between periods for a given practice, even though the initial allocation of patients to practices was genuinely random. 'Lucky' practices thus tend to become habitually profitable, and 'unlucky' ones persistently loss-making. As primary care practices are also typically very small patient

pools, the variations in profitability are likely to be very large. These characteristics suggest that the additional risk management costs of capitation contracts are likely to be substantially greater in primary care than in other health care markets, such as surgery, where demand in each period is less likely to be correlated as the same individuals do not typically return for repeat care, and where the risk pool is larger (e.g. regional hospitals).

Under capitation contracts, random risk and predictable risk are shared in exactly the same proportions. If it is deemed desirable to share a defined proportion of risk to alter provider behaviour, then the same amount of random risk is shared as well, as the two types of risk are inextricably bundled. The higher the proportion of a practice's remuneration that is fixed (that is, the higher f is and the lower v is), the higher the amount of both types of risk that is shared. Under full capitation ($v = 0$) all risks become the responsibility of the service provider, who is now the only risk-pooling entity – that is, the sole insurer. The third-party payer bears none of the financial risks of either random or predictable patient demand variation. Budget financing of a government entity is exactly equivalent to a full capitation contract, to the extent that the entity has no further recourse to government funds to 'bail it out' in the event of financial failure (the 'soft-budget constraint').

Full capitation contracts thus almost completely 'undo' the economic benefits of patients pooling random risks via large insurance companies and governments, as they fragment the original risk pool, leaving patients sharing their random demand risks and purchaser estimation errors with a very much smaller number of individuals, in pools where demand characteristics are likely to be correlated over time. The sole purpose for third-party payers under full capitation is to manage the contracting process and redistribution of resources from patients to practitioners with (presumably) lower transaction costs than the patients could achieve themselves, as the service providers become the insurers.

2.4.4 Optimal Capitation Strength

Capitation incentive strength is thus crucial for sector efficiency. Yet, determining the optimal capitation contract incentive strength (that is, setting f and v) is fraught with difficulty. The higher the degree of predictable risk shared (the higher is f , the lower is v), reducing the losses arising from over-provision, the higher the degree of random risk that is shared, raising the costs of random risk management. If too little risk is shared by the third-party purchaser (low f , high v), the desired changes in practitioner behaviour will not occur; if too much is shared (high f , low v), practice financial viability becomes a lottery, as it is largely invariant to the amount of effort practitioners put into pursuing the desired behaviour. Perverse outcomes will inevitably emerge (Holmstrom and Milgrom, 1991). As the

allocation of random risk, the propensity for each provider to engage in over-production and the propensity for each consumer to succumb to over-consumption are unique to each purchaser-practitioner pairing, each contract optimally will be unique in its setting of f and v . In practice, however, this is impossible to achieve, given the high costs of procuring information that is itself likely to be quite imperfect (Gaynor, Haas-Wilson and Vogt, 2000). Thus, it would appear that perverse consequences are inevitable under capitation, as the risks and costs are most unlikely to be well-balanced.

United States evidence suggests that substantial changes in practitioner behaviour have been achieved with very low-strength (low f , high v) incentives (Ma and Riordan, 2002). Given that United States risk pools tend to be large, this finding suggests that the losses from predictable risks, whilst real, are substantially smaller in magnitude than the benefits of aggregated random risk management. This would appear to be consistent with the United States finding that a substantial amount of variation in demand is unpredictable (Robinson, 2004; Newhouse, 1996).

In principle, therefore, primary care capitation contracts would appear to be most effective when the incentive strength is comparatively low. The most costly error would appear to be offering too strong an incentive, making practice profitability over-dependent upon luck. If the degree of predictable risk that it is desired to be controlled is high, then the amount of risk that can be shared without risking large luck-based variations in practice profitability will be greater the larger is the provider risk pool. Conversely, if the provider risk pool is small, even very small amounts of risk-sharing may tip the balance such that random risk costs rapidly overwhelm the benefits of desired changes in provider behaviour. Optimal capitation contracts will thus be highly dependent on individual practice characteristics. A 'one size suits all' approach to setting the contract incentive will inevitably mean no practice will exhibit the precise proportions of behavioural change sought. The challenge is ensuring that the incentive strength is set so that the costs of perverse outcomes do not overwhelm the size of the benefits sought from the behavioural change.

2.4.5 Comparing Health Capitation with Other Capitation Contracts

The perverse consequence of capitation contracts let by insurers 'undoing' more efficient random risk management arrangements occurs because the risk is contained within the 'closed circle' of the patient, risk manager and care provider. Both the risk manager and the service provider are agents of the patient. The patient confers the random risk management agency to the insurer, and the treatment agency to the service provider. However, risk-sharing between the insurer and the provider affects the patient directly via the treatment

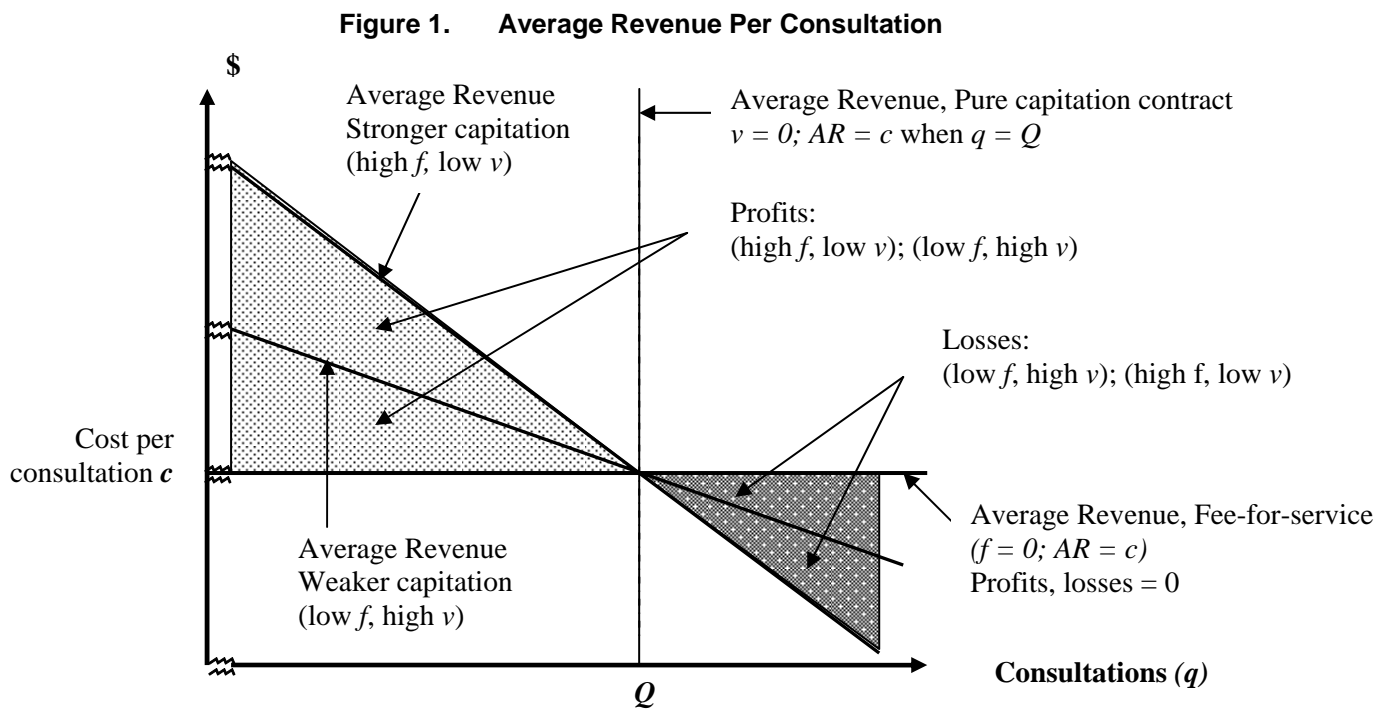
agency. This is different from ‘linear’ contractual arrangements, where risk is shared between upstream and downstream parties with distinct and separate customers, suppliers and shareholders. In this arrangement there is some room for tolerance in setting the capitation incentive strength. If, for example, the incentive contract between a shareholder and a manager is not optimal, its negative consequences can be shared by the manager with many other stakeholders, and it does not directly affect the shareholder explicitly. In the ‘closed circle’, however, there is no scope for dissipation – rather, the effects of errors are magnified as they are shafted back directly onto the party sharing random risk with the insurer in the first place.

The perversity in health care markets is exacerbated by the fact that the very reason the insurer exists is to manage risk within the closed circle. Insurer sharing of random risk necessarily reduces efficiency, even as sharing predictable risk increases it. This is different from standard linear risk-sharing (e.g. employee remuneration split between fixed and piece-rate components) as there is no other explicit risk management agreement governing the interaction. Whilst the employee contract shares predictable risk and those elements of random risk that the employer might otherwise optimally bear (e.g. business cycle variations), these risks are secondary to the purpose of the firm’s and the contract’s existence. The efficiency reduction from sharing random risk in the health case is thus of much greater consequence for industry outcomes than the random risk sharing in the employee remuneration case. This suggests that optimal incentive strengths in health care will be lower than in many other industries, in order to avoid rapid ‘unravelling’ of the benefits of insurance, further supporting the conclusion above that capitation contracts in health care markets should be sparing in the amount of risk shared with practitioners, in order to avoid patients bearing the costs of perverse outcomes.

3. *Perverse Outcomes Under Risk Imbalances*

How, then, might the effects of over-strong capitation incentives play out in primary care markets? In practice, capitation rates are often set globally for a large number of practices using information from historic service delivery averages. The values of f and v are determined assuming an ‘average’ practitioner engaging in the desired behaviours will break even financially by providing an average number of consultations Q at the prevailing average cost c . Such processes take account only of predictable risk. The aim is for all practices to deliver exactly Q consultations, and all will break even (i.e. no profits or losses). Fear of financial failure results in more costly providers seeking out ways to reduce their costs to the average level, and over-supplying practitioners reducing quantities provided. However, as

Figure 1 illustrates, the presence of random risk means that, even if practitioners take all possible steps to lower costs and eliminate unnecessary consultations, practice profits will vary. The number of consultations demanded q is now determined by random factors. ‘Lucky’ practices with low demand ($q < Q$) make windfall profits, and ‘unlucky’ practices ($q > Q$) incur losses. The stronger the contract incentive (the higher is f , the lower is v , meaning the slope of the average revenue curve is steeper), the greater the extent of the profits or losses arising from luck.



Practitioners faced with these outcomes will inevitably alter their behaviour in response to the profits and losses incurred. Bearing in mind that the objective of capitation contracts was to limit consultation numbers, any full-quality consultation in excess of Q provided by an ‘unlucky’ practitioner simply increases the size of losses incurred. ‘Unlucky’ practitioners can stem their financial losses only by limiting the number of consultations provided to Q (e.g. instituting waiting lists for consultations, ‘closing the books’ to new patients) or lowering costs of the q consultations delivered by reducing quality below the level anticipated in the contract (e.g. shorter consultation times). As the reason that the practice is ‘unlucky’ is because by dint of luck its patients are demanding more consultations than average, the patients receiving lower-quality care must, *ipso facto*, be ‘sicker than average’. The sharper the incentive (the steeper the average revenue curve in Figure 1), the greater the costs to the sickest patient pools, as the potential losses incurred by their practitioners is greater.

Conclusion 1: ‘Sicker than average’ patients in capitated systems will on average receive lower-quality service than patients who are ‘healthier than average’, simply because of the

allocation of random risk amongst small pools. The smaller the patient pools, and the sharper the contract incentive, the larger the variations in quality will be. On average the level of quality provided will decrease as the (unpredicted and unpredictable) average health state of the patients registered at a practice decreases.

Conversely, 'lucky' practices, with 'healthier than average' patients requiring fewer consultations than average ($q < Q$) face no incentive to increase the number of consultations offered above the q luck has delivered them. Each additional consultation provided above q will reduce their profits from luck (as $v < c$). Whereas under fee-for-service or price-and-volume remuneration these providers may be prepared to provide consultations to the unsatisfied patients of the 'unlucky' providers as capacity allowed, under capitation each additional consultation provided above q reduces windfall profits. There are no rewards for working hard under capitation. In fact, working harder (providing more consultations) simply reduces profits and increases the risk of financial failure. Thus, even 'lucky' capitated practices may 'close the books' to new patients, despite having potentially idle capacity (spare time) relative to the fee-for-service quantity.

Conclusion 2: The allocation of random risk to providers under standard capitation contracts reduces the supply of consultations below the average number anticipated when the random risk is borne by the purchaser. The sharper the incentives under the capitation contract, and the smaller the initial risk pools, the faster the effects of the reduction in supply will become evident, and the more rigidly the reductions in supply will be enforced by practitioners, in order to prevent financial losses or the erosion of profits gifted by 'luck'.

As capitation contracts turn service providers into insurers, the service providers now become open to the possibility of 'cream-skimming'. If service providers know about factors affecting the likelihood of an individual's likely demand for care that are not accounted for in the capitation agreement (e.g. despite exhibiting the population-based characteristics on which funding is allocated, an individual is actually healthier than average for this group and a provider has access to this information but the purchaser does not), then the provider may use this information to increase practice profits (or reduce losses) by selectively choosing to accept or reject the responsibility for insuring an individual on the basis of whether the individual's likely demand is higher or lower than the current practice average. The provider need not know about an individual's actual health state – the information about the practice where the patient was last registered, its service quality and its financial fortunes, may be sufficient to allow cream-skimming to occur.

For example, a patient wishing to transfer from a practice where quality is observably lower (due to financial distress) than the one to which the transfer is requested likely represents an increase in risk to the new practice. Quality is lower at the first practice as average health state is lower, meaning there is a high probability that the actual health state of the individual seeking to transfer is lower than the current average at the second practice. Higher demand from this individual will reduce the second practice's profits, so it may be too risky to assume the individual's risk either at all, or without further information (e.g. treating the patient on a 'casual' basis in order to establish actual health state – a practice replicating the 'stand-down' period that usually accompanies a transfer between health insurers).

Consequently, in capitated health systems, 'cream-skimming' leads to the sickest individuals being those most likely to be unable to find a health care provider, because health care and insurance are 'bundled' into the same contract. Even though the ability for a practitioner to cream-skim may be limited by imperfect prediction algorithms, a risk-averse 'lucky' practitioner can afford to take the precautionary approach of avoiding new risks without risking existing profits by simply not registering new patients. This is aided by the degree of stability in the existing patient pool – the more stable the patient pool, the greater the ability for 'lucky' practitioners to maintain profitability by 'closing the books'. Thus, cream-skimming need not be a deliberate activity based upon specific knowledge about a patient – rather, not registering is the dominant strategy for any 'lucky' practice with a stable patient list. The effect of both types of cream-skimming is the same – a separation of patients between practices, based upon health care demand, simply because of the random allocation process.

Conclusion 3: Cream-skimming of patients will necessarily be associated with capitation contracts, simply because of the systemic effects of creating habitually lucky practices via random allocation. Deliberate cream-skimming using private information occurs in addition to this systemic effect. Again, the sharper the capitation contract incentive, and the smaller the initial risk pools, the more aggressive the cream-skimming undertaken will be, and the greater the distortion in the allocation of patients by health state amongst practices.

The 'flip side' of the incentive for 'luckier' practitioners to benefit from 'cream-skimming' is the propensity for 'unlucky' practices to grow larger in order to reduce the effects of random risk. Whereas a 'lucky' practice faces the risk that a new patient is most probably 'sicker than the current practice average', the probability is greater than 0.5 that any new patient requesting registration at a 'sicker than average' practice is actually healthier than the current practice average (as the total population average, minus the 'sicker than average' practice,

will be healthier than the population average including the ‘unlucky’ practice). There is a better than even chance that the new patient will raise the average health state of the practice. By this reasoning, in a capitated system, only ‘unlucky’ practices will willingly register patients of unknown health state. Consequently, ‘unlucky’ practices will grow larger, replicating many of the benefits of spreading risk via a large risk pool. However, these practices will still be ‘unluckier’ than the average practice, as the ‘lucky’ practices have skimmed off more than the average number of healthier-than-average individuals. Growing larger simply minimises the extent of the losses to which the ‘unlucky’ practices are exposed.

On the one hand, the benefits of size suggest that the pressure of financial failure amongst ‘unlucky’ practices might lead to the active pursuit of mergers⁴ with other practices in order to stem the financial losses. On the other hand, no rational practice owner will want to merge with a practice that is clearly ‘unluckier’ than his (and by the mere fact of seeking to merge, a practice may be signalling that it is ‘unlucky’). If a practice is making profits, merging with the ‘unluckier’ practice (even if ‘lucky’ in absolute terms in that it is making positive profits) means that profits must be shared, and the ‘luckier’ practitioner receives less average profit per consultation under merger than under separation. Likewise, no loss-making practice owner will want to merge with a practice whose losses are even larger, as the first practice owner will end up worse-off, becoming responsible for larger average losses in the combined entity relative to remaining separate.

In practice, the problem of increasing practice size to reduce risk costs is solved by bifurcation of ownership form. A private owner of a loss-making practice must personally bear the costs of losses from bad luck out of his own reserves (either cash injections, or accepting a very much lower return on his human capital than that received by ‘lucky’ practitioners). If these costs are intolerable to the owner, or better opportunities are present (e.g. being hired as a salaried employee, where there is no obligation for a practitioner to bear any financial risk personally, as returns for effort under salary are certain), then the practitioner-owner of an ‘unlucky’ practice is better off ‘walking away’ than retaining ownership. But, as the practice is loss-making, no rational owner will be prepared to invest in it. A missing market for loss-making practices emerges.

The only ownership group likely to voluntarily take up responsibility for serving the loss-making pools will be ‘non-owned’ nonprofit and charity providers, who can make up the

⁴ The use of ‘merger’ here means the merger of patient lists in a truly joint practice where the providers become equity-sharing partners – that is, the substitution of sole ownership with a firm (i.e. ‘corporate’ form). This is distinct from the co-operative form of independent practice associations or ‘group practice’ where each practitioner retains his own list.

deficits from other means not available to private owners (e.g. donations, tax exemptions), or even government (via taxation). Such providers will likely be less averse to the prospect of merging practices and spreading the losses amongst a very much larger pool. Whilst such actions will reduce the total costs of risk management, the large, ‘unlucky’ pools will still on average be making losses on the contract relative to the risk-free state, as ‘lucky’ practices are still extracting profits from their healthier than average patient pools. Moreover, if the additional sources of revenues available to the non-owned providers are insufficient to cover the losses incurred, quality offered in the non-owned practices will still be lower than that in the privately-owned, ‘lucky’ practices. Ironically, many nonprofit and charity providers often end up re-hiring as salaried employees the very practitioners who have ‘walked away’ from the practice because of its financial failure.

By contrast, an aggressive market will emerge for private ownership of the ‘lucky’ practices. The ‘luckier’ (more profitable) the practice is, the greater the ‘bonus’ to the first owner under capitation contracts. This owner not only makes and extracts large profits from luck, but when selling the practice, will be able to charge a premium (‘goodwill’) in proportion to the degree of luck enjoyed. Prospective owners will be prepared to pay more for a more profitable practice up to the point where the premium paid equals the long-term profits able to be earned from luck. However, only the first owner enjoys this bonus. The purchase price premium paid by the second and subsequent owners increases the costs of service provision. Second and subsequent owners need to recoup not just a fair return on time expended, but also a fair return on the additional capital supplied, which was not present under risk-free contracts. Indeed, the need for larger amounts of physical capital to purchase such practices may be a factor in inducing private sector non-practitioner ownership of primary care practices in capitated systems, in addition to the promise of gains arising from skilled insurer-investors arbitraging against practitioner-owners’ comparative informational disadvantages in managing risk pools.

Conclusion 4: Capitation unequivocally raises the costs of care delivery, not just by the additional payments injected to cover losses by ‘unlucky’ practices (both private and charity sources) but also by the introduction of capital costs in the financial structure of lucky practices not present under fee-for-service. The sharper the capitation incentives, and the smaller the initial risk pools, the greater the additional costs incurred relative to fee-for-service.

Conclusion 5: Allocation of random risk in capitation contracts to practices will likely lead to changes in the patterns of institutional size and ownership form in primary care markets. A

bifurcation between small 'lucky', profitable practitioner-owned practices and large, 'unlucky' unprofitable practices served by nonprofit and charity providers with practitioners as salaried employees will emerge. The sharper the capitation incentive, and the smaller the initial risk pools, the faster this is likely to occur.

Conclusion 5a: By combining Conclusions 1 and 5, the bifurcation of size and ownership on the basis of practice 'luck' will be further underpinned by quality differences. If nonprofit and charity owners are unable to cover all losses from additional sources, the bifurcated market will be characterised by 'unlucky' large, nonprofit and charity providers, whose patients are on average less healthy, and who will likely receive lower quality care, and small, practitioner-owned providers whose patients are healthier than average and receive higher-quality care than the patients of the non-owned practices. Once again, the sharper the capitation incentive, and the smaller the initial risk pools, the faster the two-tier system will become established, and the greater the quality differences will be.

Conclusions 1, 2 5 and 5a lead to a further conclusion that capitated systems will also be associated with a substantial degree of practitioner self-selection. Capitation contracts provide no rewards for working hard (Robinson, 2001). Practitioners who do not care to work as hard will prefer to become owners of the small, privately-owned practices where higher profits are earned for less effort. However, if human capital is increased by experience gained from delivering more consultations and delivering those consultations to individuals who are 'sicker than average' (McGuire, 2000), then human capital will be accumulated faster by practitioners in 'unlucky' practices.

On the one hand, higher levels of human capital may partially offset some of the costs of lower quality borne by patients of 'unlucky' practices, as per conclusion 1. On the other hand, it suggests a second self-selection (adverse selection) problem. The practitioners with the greatest motivation to work hard to build human capital tend to be young and comparatively inexperienced. They also tend to lack the capital to buy the now highly-valued 'lucky' practices where profits can be made for exerting less effort. These practitioners will likely select, in disproportionately large numbers, to work (as salaried employees) in 'unlucky' practices. Thus, in absolute terms, the difference in actual human capital employed may not be very great, especially if the younger practitioners tend to work hard initially to accumulate experience and the resources to subsequently fund the purchase of a private practice, which allows them to work at a more leisurely pace in the latter part of their careers.

Thus, ‘unlucky’ practices will tend to be serviced by younger less-experienced but more hard-working practitioners than the ‘lucky’ practices⁵.

The adverse selection problem, however, becomes more acute when hard-working practitioners of all ages and any amounts of human capital can select between capitated systems and fee-for-service ones. Fee-for-service systems provide no penalties for working harder than average – rather, such behaviour is rewarded. Thus, hard-working practitioners will accrue greater returns (both in accumulating human capital and generating a greater return investments in both human and physical capital) by working in fee-for-service systems than in capitated ones. This suggests that where choice of contract type exists, adverse selection will affect the type of practitioners opting to practice in each system.

Conclusion 6: Where choices of contract type exist, hard-working practitioners and those seeking to increase human capital more rapidly will self-select into contracts that minimise the amount of risk shared with practices (fee-for-service) rather than into risk-sharing ones (capitation, price-and-volume). The long-run effect of such self-selection will be that when there is competition between different contract types, the average practitioner remunerated by the capitated contract will be less experienced and less hard-working than the average practitioner remunerated by a fee-for-service contract. Specifically, young practitioners seeking to accumulate human capital quickly will select away from markets remunerated by capitation contracts. Whilst some looking for an ‘easier life’ may buy into capitation practices later, thereby raising average human capital, those who genuinely wish to be well-remunerated for working hard will eschew the capitated system. Once again, the sharper the capitation incentive, the faster the effects of self-selection will emerge, and the more costly they will be.

Ultimately, the effects of an improperly-balanced capitation contract play out in detriments to patients, in the form of either higher direct costs of primary health care (e.g. higher premiums, taxes or patient co-payments) or indirect costs associated with service quality than if the contract had been better-balanced. The balance is fine, and as this section illustrates, the consequences of getting the balance wrong may be very large.

⁵ This finding is consistent with, but derived from a different origin to, the altruism literature that suggests practitioners who are more altruistic will self-select into the sectors where lower-quality care is delivered, in order to ‘even up’ the distribution of available care – see, for example, Delfgaauw, 2007; Glazer, 2002; Ma, 2004.

4. Policy Implications

The preceding sections lead to the conclusion that the use of financial risk-shifting in health care contracts carries with it additional effects not normally considered in the agency theory of contractual risk-sharing, as a consequence of the fact that the purchasing party sharing the risk is already itself undertaking a risk management role. Clearly, capitation contracts have long-term effects upon industry ownership and structure that are not costless, and must be considered when evaluating the efficacy of using a capitation instrument. Moreover, these effects will be different in health care markets to those expected in other sectors, simply because the types of risks being shared are different, and more acute in primary care markets due to the historic patterns of small, practitioner-owned businesses and repeat transactions between the same patients and practitioners leading to correlations in risks across time.

These factors tend towards the conclusion that capitation contracts are less desirable in primary care markets than other health care markets, and when applied the optimal incentive strength must be quite small in order to avoid the costly consequences of reallocation of random risks crowding out the benefits of predictable risks. The six conclusions above suggest that the costs of getting the incentive strength wrong are substantial not just in respect of the costs of care, but also in the allocation of care and in respect of the choices made by practitioners when investing their human capital. Unless the incentives are carefully balanced, substantial changes in sector organisation are likely to emerge, and these changes may not necessarily be in accordance with the objectives underpinning their implementation.

In particular, by their handing of random risk, primary care capitation contracts lead to allocation systems that are less equitable in the allocation of the available resources on the basis of health need than either fee-for-service or price-and-volume contracts. It is thus flawed logic where capitation contracts have been implemented as a means of engendering greater equity in the allocation of primary care resources (although it is noted that this may not be as strong an argument in respect of other types of health care, where correlated demand is less of a concern). They are more equitable only if individual patient demand for primary health care is perfectly predictable and the propensity for both patient- and practitioner-induced over-consumption is costlessly ascertainable, enabling a perfect incentive to be determined for each practice. Moreover, the incentive strength must be adjusted each time a patient or practitioner either leaves or joins the practice.

Clearly, none of the conditions for optimal capitation prevail. Rather, as it is very costly, if not impossible, to ascertain the necessary information to set perfect incentives, and most

systems adopt a ‘one incentive strength for all’ approach (given differences for patients of different classes ex ante), the incentive will inevitably be ‘wrong’. If the incentive is too soft, the desired changes will not be achieved. If it is too strong, the result will likely be a substantially inequitable bifurcated system that is ‘fair’ to neither patients nor practitioners, and which costs substantially more per consultation than a fee-for-service system delivering the same number of consultations. This suggests that a capitated system will be of benefit only where there is evidence of substantial existing distortions from over-consumption, and then only if the incentive strength used is low. Given that the optimal strength for primary care capitation contracts is low, and the costs and likelihood of getting the strength wrong are substantial, unless there is evidence of such distortions, then it may be easier and more cost-effective to forego the use of capitation contracts in primary care, and concentrate instead on the use of non-financial instruments to achieve the desired behavioural changes.

In the paper from which the opening quote in this discussion is taken, James Robinson muses why, given the extensive use of risk-sharing agreements in other sectors, use of the piece rate (i.e. payment per consultation) has persisted in health care. He concludes that there is no ideal compensation contract for physicians. He advocates a mixture of payment methods that minimise the costs of the side-effects of each type of contract. As a general principle in health care contracting, this approach appears reasonable. However, this paper would place a caveat on Robinson’s recommendation in suggesting capitation may be so fraught with difficulty in primary care settings that it should be employed as a last, rather than a first, resort when changes in practitioner behaviour are sought, and certainly not as an instrument to allocate system resources more ‘fairly’ based on health need or practitioner reward for effort in respect of providing consultations on the basis of health need. It would also go so far as to suggest that, in respect of the allocation of random risk in primary care, the piece-rate has persisted as it is the least distorting payment method, and provides the greatest incentives to practitioners to work hard to improve the health outcomes for their patients. Policy-makers and insurers would do well to bear this in mind when contemplating using their market power to unilaterally impose capitation contracts in primary care markets.

References

- Arrow, Kenneth. 1963. Uncertainty and the welfare economics of medical care. *American Economic Review* 53(5): 941-73.
- Boyd, Haleigh; Lewis Evans and Neil Quigley. 2000. The efficiency of private agricultural contracts. *Victoria University of Wellington Law Review* 31: 813-846.
- Coase, Ronald. The nature of the firm. *Economica* 4.
- Cooper, David J.; and James B. Rebitzer. 2006. Managed care and physician incentives: the effects of competition on the cost and quality of care. *B.E. Journals in Economic Analysis and Policy: Contributions to Economic Analysis and Policy* 5(1): 1-30.
- Coster, Gregor; and Barry Gribben. 1999. *Primary Care Models for Delivering Population-based Health Outcome*. Wellington, New Zealand: National Health Committee.
- Crampton, Peter; Frances Sutton and Jon Foley. 2001. Capitation funding of primary care services: principles and prospects. *New Zealand Medical Journal* 115 (1155).
- Cumming, Jackie. 1999. *Funding Population-Based Primary Care in New Zealand*. Wellington, New Zealand: National Health Committee.
- Cutler, David; and Richard J. Zeckhauser. 2000. The anatomy of health insurance. Chapter 11 in Culyer, A and Newhouse, J (eds). *Handbook of Health Economics*. Amsterdam: Elsevier.
- Danzon, Patricia M. 1997. Tort Liability: a minefield for managed care. *Journal of Legal Studies* 26: 491-519.
- Delfgaauw, Josse. 2007. *Dedicated Doctors: Public and Private Provision of Health Care with Altruistic Physicians*. Tinbergen Institute Discussion Paper 2007-010/1. Available on <http://www.ssrn.com/abstract=958693>.
- Dranove, David; and Mark Satterthwaite. 2000. The Industrial Organisation of health care markets. Chapter 20 in Culyer, A and Newhouse, J (eds). *Handbook of Health Economics*. Amsterdam: Elsevier.
- Dranove, David; Carol Simon and William White. 2002. Is managed care leading to consolidation in health care markets? *Health Services Research* 37(3): 551-571.
- Dranove, David; Carol Simon and William White. 1998. Determinants of managed care penetration. *Journal of Health Economics* 17: 729-45.
- Ellis, Randall; and Thomas McGuire. 1986. Provider behaviour under prospective reimbursement: cost-sharing and supply. *Journal of Health Economics* 5(2): 129-51.
- Gaynor, Martin, Deborah Haas-Wilson and William Vogt. 2000. Are invisible hands good hands? Moral hazard, competition, and the second-best in health care markets. *Journal of Political Economy* 108(5): 992-1005.
- Glazer, A. 2004. Motivating devoted workers. *International Journal of Industrial Organization* 22(3): 427-440.
- Hansmann, Henry. 1996. *The Ownership of Enterprise*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.
- Hagen, Gary. 1999. Shaky Ground: questioning health plan and provider solvency. *Risk Management* 46(11): 42-45.
- Hart, Oliver D.; and Sanford J Grossman. 1986. The costs and benefits of ownership: a theory of lateral and vertical integration. *Journal of Political Economy* 94(4): 175-202.
- Holmstrom, Bengt; and Paul Milgrom. 1991. Multi-task principal-agent analyses: incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organisation* 7: 24-52.
- Honore, Peggy A.; Eduardo J. Simones; Ramal Moonesinghe; Harold C. Kirbey and Meg Renner. 2004. Applying principles for outcomes-based contracting in a public health programme. *Journal of Public Health Management and Practice* 10(5): 451-457.
- Howell, Bronwyn. 2006. *Unveiling Invisible Hands: two-sided platforms in health care markets*. Paper presented at the European Association for Research in Industrial Economics conference, Amsterdam, August 25-27, 2006.

- Howell, Bronwyn. 2005. *Restructuring primary health care markets in New Zealand: from welfare benefits to insurance markets*. Australia and New Zealand Health Policy 2:20. Available on <http://www.anzhealthpolicy.com/content/2/1/20>.
- Keen, Nicholas; Donald Light and Nicholas Mays. 1999. *Public-Private Relations in Health Care*. London, United Kingdom: King's Fund.
- King, Annette. 2001. *The Primary Health Care Strategy*. Ministry of Health. Available online [http://www.moh.govt.nz/moh.nsf/0/7BAFAD2531E04D92CC2569E600013D04/\\$File/PHCStrat.pdf](http://www.moh.govt.nz/moh.nsf/0/7BAFAD2531E04D92CC2569E600013D04/$File/PHCStrat.pdf).
- Ma, Ching-To Albert. 2004. *Altruism and Incentives in Public and Private Health Care*. Mimeo, Boston University.
- Ma, Ching-To Albert; and Thomas G. McGuire. 2002. Network incentives in managed health care. *Journal of Economics and Management Strategy* 11(1): 1-35.
- Ma, Ching-To Albert; and Michael H. Riordan. 2002. Health Insurance, Moral Hazard and Managed Care. *Journal of Economics and Management Strategy* 11(1): 81-107.
- Malcolm, Lawrence. 1997. GP budget-holding in New Zealand: lessons for Britain and elsewhere? *British Medical Journal* 314(7098): 1890-2.
- Martin, Lawrence L. 2002. Performance-based contracting for human services: lessons for public procurement. *Journal of Public Procurement* 2(1): 55-71.
- McGuire, Thomas G. 2000. Physician Agency. Chapter in Culyer, A and Newhouse, J (eds). *Handbook of Health Economics*. Amsterdam: Elsevier.
- Milgrom, Paul; and John Roberts. 1992. *Economics, Organization and Management*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Newhouse, Joseph P. 1996. Reimbursing health plans and health providers: efficiency in production versus selection. *Journal of Economic Literature* 34(3):1236-1263.
- Newhouse, Joseph P. 1973. The economics of group practice. *Journal of Human Resources* 8(1): 37-56.
- Pauly, Mark V. 2000. Insurance Reimbursement. Chapter 10 in Culyer, A and Newhouse, J (eds). *Handbook of Health Economics*. Amsterdam: Elsevier.
- Pauly, Mark V. 1968. The economics of moral hazard: comment. *American Economic Review* 58: 531-37.
- Rivers, Patrick Asubonteng; and Kai-Li Tsai. 2003. Managing costs and managing care. *International Journal of Health Care Quality Assurance* 14(6/7):302-307.
- Robinson, James C. 2004. Reinvention of health insurance in the consumer era. *Journal of the American Medical Association* 291(15): 1880-1886.
- Robinson, James C. 2001. Theory and practice in the design of physician payment incentives. *The Milbank Quarterly* 79(2): 149-177.
- Scott, Anthony. 2000. Economics of General Practice. Chapter 22 in Culyer, A and Newhouse, J (eds). *Handbook of Health Economics*. Amsterdam: Elsevier.
- Williamson, Oliver. 1986. *Economic organization: firms, markets and policy control*. Brighton, Sussex: Wheatsheaf.
- Zeckhauser, R. 1970. Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. *Journal of Economic Theory* 2(1): 10-26.
- Zweifel, Peter; and Willard G Manning. 2000. Moral hazard and incentives in health care. Chapter 8 in Culyer, A and Newhouse, J (eds). *Handbook of Health Economics*. Amsterdam: Elsevier.