Health Insurance Reform in Four Latin American Countries: Theory and Practice

William Jack

1This paper was prepared as part of a regional analysis of social risk management in the Latin America and Caribbean region of the World Bank. Thanks, without implication, to Girindre Beeharry, Chico Ferreira, Charlie Green, Je Hammer, Jeni Klugman, Jerry LaForgia, Maureen Lewis, Andre Medici, and Miryan Sato for discussions. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent. The paper should not be cited without the permission of the authors.

2Department of Economics, Georgetown University, Washington DC.
Abstract

This paper first examines rationales for public intervention in health insurance markets from the perspective of public economics. It then draws on the literature of organizational design to examine in some detail alternative public intervention strategies, including issues of contracting, purchaser provider splits, and regulation of competition. Health insurance reforms in four Latin American countries are then considered in light of the insights provided by the theoretical literature.
1 Introduction

Health care expenses and lost labor earnings due to illness represent a major source of risk for individuals and families. Exposure to such risks is costly in itself (if individuals are risk averse), but can also have long-term effects especially on the poor. Selling assets, withdrawing children from school to care for ill parents, and exiting the labor market can leave low-income families trapped in poverty.

Protection against the risks of ill health can be achieved by reducing the size and variability of the underlying stochastic process, for instance by improving public goods that affect health outcomes (pollution, etc.), and by spreading risks across individuals. This paper will address the second of these risk reduction strategies, with particular emphasis on the design and organization of the relevant institutions in Latin America.

Given its income, Latin America is on average a relatively healthy part of the world. For example, Figure 1 shows data from 157 countries that reported information on GDP per capita and life expectancy in 1997 (Latin American countries are represented as squares, the others as diamonds\(^1\)).

Visual inspection of the simple logarithmic regression indicates that most Latin American countries have life expectancies at birth equal to or significantly above those predicted by their income levels.

Health outcomes measured by infant mortality rates are less positive however, as shown in Figure 2. Similarly, the distribution of health care services within Latin American countries is skewed to the better off. Recent analysis undertaken by PAHO, UNDP,\(^2\) and the World Bank, (Greene et al. (1999)) indicates wide disparities in both health needs and access to medical care across income groups. This, coupled with incomplete and fragmented insurance coverage, led several countries in the region to adopt wide-ranging health sector reforms in the latter part of the 1980s and the 1990s that continue today.

Large scale changes in health insurance and health care markets inevitably involve significant

\(^2\)Pan American Health Organization, and United Nations Development Program, respectively.
Figure 1:
Figure 2:
public intervention. The purpose of this paper is first, in section 2, to examine the arguments in favor of such intervention from a public economics perspective. Having identified market failure and redistributional rationales for public intervention, section 3 addresses the important issue of how the government should intervene. This is effectively a question of organizational design, incorporating ideas from industrial organization, contract theory, and the theory of the firm. In this light, section 4 undertakes a detailed examination of the reforms pursued in Colombia, Argentina, Brazil, and Chile. These countries reflect a variety of routes toward the goals of expanding formal insurance coverage and improving the efficiency of service delivery in their health systems. These directions of reform are interpreted within the framework developed in sections 2 and 3. Section 5 concludes.

2 Reasons for government intervention in the health sector

The theoretical literature on the performance of insurance markets is well developed. Not all of the market failures that may arise in such markets, however, necessarily justify public intervention. This section examines the efficiency and equity reasons for intervention in health insurance markets, paying specific attention to the informational constraints facing governments.

2.1 Market failure in the health insurance sector

It is useful to briefly review sources of inefficiency in the delivery of health insurance, and to examine the extent to which public intervention can correct the associated market failures. The inefficiencies derive mainly from information asymmetries and imperfect competition, and less from standard public goods and externality characteristics.

Asymmetric information — moral hazard

The role of information in the performance of insurance markets has been widely appreciated. In the health insurance literature, Feldstein (1973), Pauly (1968), and Zeckhauser (1970) showed how asymmetric information at the ex post stage — that is, after an insured event has occurred — can
reduce the efficiency of equilibrium insurance contracts. This moral hazard occurs when insurance contracts are written on the basis of endogenous incurred expenses and not on the basis of exogenous health needs. This kind of insurance leads to over-consumption of care, the distortionary costs of which are offset by reducing the level of insurance.

A similar inefficiency results from “ex interim” moral hazard, when precautionary actions can be taken after the insurance contract is signed, but before uncertainty is resolved. In this case, the inability to make insurance contracts contingent on such actions reduces the optimal level of insurance. In both cases, the individual is second-best optimally exposed to some risk.

Within a partial equilibrium model neither source of moral hazard argues for public intervention, unless one assumes unrealistically that the government has better information than private insurers. On the other hand, Greenwald and Stiglitz (1986) showed that, taking a general equilibrium approach, there may be a role for government intervention, even when the government does not have an information advantage vis-à-vis private insurers. Their argument is simply that, through its powers of taxation and subsidization, the government can encourage desirable ex interim actions by altering the prices of goods and services that have non-zero cross elasticities of demand with such actions. Thus, taxing cigarettes may reduce the (assumed unobservable) ex interim action of smoking, thus mirroring an efficient insurance contract. The scope for such Pareto improving intervention becomes more limited as private insurance contracts become more sophisticated, e.g., by disallowing benefits to smokers.

**Asymmetric information — adverse selection**

While moral hazard derives from asymmetric information that is generated after contracts are signed, adverse selection occurs in markets when information is held asymmetrically at the date of contracting. A competitive insurance market in a population with heterogeneous ex ante risk characteristics may perform inefficiently if insurance contracts cannot be differentiated on the basis of these risks.

When only a single insurance policy is available, Akerlof’s (1970) lemons problem may emerge,
with a proportion of individuals choosing not to purchase insurance. On the other hand, if multiple contracts are available, then even when risk characteristics are unobservable it becomes possible for insurers to charge low-risk individuals lower prices. All individuals will have some insurance in an equilibrium (Rothschild and Stiglitz (1976)), but two problems may arise: first, the good risks will not have full insurance, and second, an equilibrium may not exist.

Evidence of the lemons-type of equilibrium (in which a fraction of the population is uninsured) has been provided in a number of studies (e.g., Cutler and Zeckhauser (1997), and Cutler and Reber (1998)). The relevance of the Rothschild-Stiglitz equilibrium is possibly more debatable, since it is typically high-risk individuals who have trouble obtaining full insurance against health risks, and not low risks, as their model suggests. However, since risk is correlated with other determinants of insurance coverage (such as income, education, etc.) it is probably imprudent to dismiss the underlying model.

Government intervention cannot easily correct these market failures. In both models, universal and uniform coverage can be mandated, but the resulting resource and risk allocations are not Pareto-comparable with the initial equilibrium.

**Imperfect competition**

The models of adverse selection reviewed above identified failures of competitive insurance markets. But even in the absence of adverse selection problems, insurance markets may yield socially sub-optimal resource and risk allocations if firms have market power. Such market power may derive from information imperfections on the demand side, say contributing to switching costs (which make it difficult for new firms to attract customers). Also, increasing returns in administrative costs suggest that a somewhat concentrated industry is likely to be observed in equilibrium.

In standard industrial organization models, while market power typically leads to allocative inefficiency, what competition there is is generally welfare improving. However, in insurance markets with information asymmetries, competition may sometimes have negative effects on allocative efficiency. For example, when insurers are faced with a heterogeneous risk population they will have
incentives to sell policies only to low-risk individuals — i.e., those individuals to whom it is cheap to provide insurance. If they cannot offer different policies to different risk types, then they may lower the quality of the policies they do sell to make them sufficiently unattractive to high-risk individuals (Jack, forthcoming). This kind of selection incentive might suggest public intervention to control the extent, or at least type, of competition in the insurance market. I will refer to this effect as active selection — the selection by firms, because of their policies, of good risks, as opposed to adverse selection — the selection by firms, in spite of their policies, of bad risks.

**Consumer protection**

*Quality of care*

The examples of moral hazard above concentrated on the behavior of consumers when they are at an informational advantage vis-à-vis the provider of insurance. One response by insurers is to try to improve the information they have about consumers, by undertaking “utilization reviews” — essentially checking that doctors are not providing “too much” care. In order to motivate doctors, insurers may indeed give them stakes in the insurance company, converting it to a managed care organization.

Such an organizational design is efficient, as long as the information asymmetry is removed — that is, as long as information about health status (and the effects of medical interventions and their costs) is held symmetrically by all parties. In practice the physician is the primary source of this information, so that when acting as the patient’s agent he confers an information advantage on the patient vis-à-vis the insurer. However, when acting for the insurer, the physician may put the patient at a disadvantage, and warranted treatments could be withheld. The usual competitive forces that induce firms to keep quality high may not work well in this situation, and quality of care could suffer. One can appeal to the literature on consumer protection and safety standards in support of the role of government in markets with uninformed consumers.

*Financial regulation*

Insurance companies perform similar functions to banks. Banks facilitate inter-temporal trades
(saving and dissaving) implemented through contemporary inter-personal trades (lending and borrowing), while insurance companies facilitate trades between uncertain states of nature implemented through inter-personal pooling of current risks.

The transformation function of banks (converting short-term savings into long-term project funds) results in a mis-match between the time profiles of assets and liabilities in banks’ balance sheets, introducing a moral hazard problem with regard to the behavior of asset managers. Similarly, insurance companies tend to collect premiums in advance of the resolution of uncertainty, and may optimally build up resources in order to self-insure against systemic risks. In this case, insurance company managers must choose how and where to invest these funds. A feature of health insurance that is not common to banks in this regard is the fact that the real value of the firm’s assets is a function of two factors: the quality of financial investments, and the costs of providing medical services.

Dewatripont and Tirole (1994) use an incomplete contracts model to show that bank managers can be given appropriate incentives to perform by transferring control from equity holders (who have relatively weak incentives to interfere with management) to debt holders (whose incentives to interfere are stronger) when bank performance as measured by the value of assets is poor. This allocation of authority is in place of a more sophisticated, but infeasible, explicit performance contract. The role of government is then to act as a representative of small, uncoordinated debtors, and the theory rationalizes public take-overs of distressed banks as a means of providing incentives to managers.

A similar role can be ascribed to the government with respect to health insurance regulation. If the value of a firm’s assets falls enough, the government may wish to intervene and take over the administrative functions (maybe contracting out such administration to another healthy insurer).

**Life-time insurance and non-diversifiable risks**

A final important failure of insurance markets is that they often do not provide life-time insurance. Since individuals’ health needs exhibit a degree of autocorrelation, insurance that is actuarially fair only on an annual basis exposes the individual to a high variance of medical costs over the lifetime.
Part of the reason it is difficult to provide life-time insurance is that the future development of medical care prices is itself uncertain, due mainly to the vagaries of technological and epidemiological dynamics. These components of risk are systemic, so cannot easily be insured against (except inter-temporally, across generations). The government faces the same constraints as private insurers in terms of the nature of the systemic risks, but may be in a better position to facilitate inter-generational trades due to its power of taxation.

2.2 Equity

Efficiency is a relatively uncontroversial goal of economic policy and organization. However, in the health care and health insurance sectors, equity is equally, if not more, important in shaping policy. This may be because individuals are more willing to accept differences in general income levels than in the status of health across individuals.

At a conceptual level, having high medical needs, or being at high risk of needing medical attention, reduces an individual’s available budget set. In consequence, the government may wish to redistribute resources between individuals with identical money incomes. In practice, the redistribution from healthy to sick is often attempted by imposing uniform prices for health services and for health insurance across individuals. Of course, charging uniform prices for doctor visits is not redistributive if the sick must visit the doctor more often than the healthy.

On the other hand, in insurance markets, uniform pricing of insurance policies across individuals is a feasible tool for redistribution from low risks to high risks. However, requiring private firms to community rate may only exacerbate adverse selection and active selection problems that already exist.

It is important to note that, even if selection issues were unimportant, it would still be only second-best to require uniform insurance pricing. The first-best policy, of course, is to redistribute income (lump-sum) from low risks to high risks, and to require each individual to buy insurance at the actuarially fair price (i.e., to allow price discrimination by firms). Such redistribution is notoriously
difficult, and even more so when income inequality itself is high, as it is in many Latin American countries. Indeed, a social welfare maximizing planner will likely wish to redistribute from rich to poor, and from less risky to more risky. This multi-dimensional problem of redistributive taxation is difficult, even when the government restricts itself to simple linear tax and transfer systems (Jack (1999)).

Henriet and Rochet (1999) have recently analyzed the optimality of a uniform public insurance system within the context of a similar multi-dimensional redistribution problem. In the absence of moral hazard, they find that a comprehensive insurance policy providing full insurance to all individuals is part of an optimal tax and insurance system. This result relies to some extent on the assumption that individuals face the same distribution of losses, albeit with different probabilities. As the demand for health care is income elastic, however, the financial cost of care consumed when ill is an endogenous function of income (and is in fact limited by the individual’s available income). This suggests that the poor may prefer additional income transfers and less extensive public insurance to being offered the same level of coverage as the non-poor. Alternatively, the rich may be willing to pay somewhat higher taxes in order to have more comprehensive insurance than the poor.

This discussion leads us to expect that in the presence of health risk and income differentials across the population, and in the absence of first-best redistributive taxation, governments will likely wish to couple a progressive general tax source (e.g., an income tax) with a system of health insurance (privately or publicly supplied — more on this later) that delivers subsidized insurance to the poor, but allows coverage to increase with income.

One means of effecting such a graduated insurance profile is to have a mixed public/private system of insurance, in which the government provides (or mandates) a given base level of insurance, and individuals are permitted to top up their coverage through private purchases, or to opt out of the public system and purchase private insurance. For example, Besley and Coate (1991) have shown that public provision of insurance (of low enough quality) can be used as a redistributive tool as long as individuals have access to supplementary private coverage. Using a political economic model
with three classes of voters, Gouveia (1997) also establishes the use of opting out as a means of implementing non-uniform insurance coverage in the presence of a progressive income tax. Such a system is supported by the rich, since the cost of purchasing their preferred level of (privately provided) insurance is less that the taxes they would pay for a similar level financed through taxation. The poor also support the mixed system because they have a lower demand for coverage — they would prefer to save some of their taxes that would finance better coverage in order to purchase other goods.  

3 The nature of government intervention

The arguments above have suggested that either due to market inefficiencies (mainly adverse selection) and redistributive concerns, governments may wish to control individuals’ choices about insurance in certain ways. What the discussion does not tell us is how such control over choices should be effected. In this section we examine the design of public interventions recognizing that the productive efficiency of insurance is a function of the administrative costs incurred and the costs of providing covered services. Costs of provision of insurance and medical care can be controlled through explicit contractual arrangements between insurer and provider, organizational choices (e.g., vertical integration), and competition. These three avenues are considered in turn.

3.1 Motivating physicians

In many markets, providers of goods or services are paid on the basis of outputs, not inputs, providing strong incentives for efficient production. It is well understood however that insured medical services

\[3\text{There is a growing literature on normative and political economic models of the public provision of private goods - see Blomqvist and Christiansen (1999) for a useful summary. Uniform public provision of a private good, such as health insurance, can be welfare improving to the extent that imposing quantity constraints on some individuals choices weakens self-selection constraints in the optimal income tax problem. These ideas are further elaborated on in Blomqvist and Christiansen (1998a) and (1998b).}\]
are nearly never compensated in this fashion. For example, an insurance policy that reimburses incurred costs gives the provider no incentive to choose an efficient input mix.

The underlying problem here is one of motivating physicians and hospital administrators. When neither health output nor physician effort is directly purchasable (i.e., contractible) then effort incentives may be difficult to generate without exposing the physician to undesirable risk. At two extremes are the salaried physician and the decentralized fund holder. Under the first arrangement, an insurer (possibly the government) pays the physician a fixed amount, independent of his supply of effort, and reimburses non-physician expenses (e.g., laboratory tests). In the absence of ethical concerns or job satisfaction (which are likely to be significant in many instances), doctors will tend to substitute out of personal effort and into complementary inputs. On the other hand, a decentralized fund holder, who is given a fixed budget to finance all incurred costs (including the cost of his own effort) will have strong incentives to choose the right input mix.

The trade-off, as in any moral hazard problem, is between incentives and risk. A salaried physician faces little personal financial risk associated with expensive cases, since he is reimbursed for other costs. On the other hand, a fund holder with good effort incentives must bear the full costs of treating expensive cases, and so might be exposed to considerable risk. This generates a potential alternative source of inefficiency: if providers attempt to reduce their risk exposure by inducing hard-to-treat patients to switch physicians, then these active selection attempts can increase equilibrium costs, or reduce quality. If instead the physician is provided with some insurance against hard-to-treat cases directly by the payer, then his incentives to perform are weakened.

It should be clear from this discussion that the trade-offs between risk and incentives at the provider level are not specific to the public sector. Private insurance companies also must induce physicians to implement insurance contracts at minimum cost, and so face a similar optimization problem. Some endogenous differences across the public and private sectors may emerge however that imply different contractual relationships between payers (the government, or insurance companies) and providers of medical care. For example, if public insurance is aimed at the poor, then in the
absence of accurate eligibility tests self-selection constraints may require that the public system provide a relatively low quality of service. Inducing low effort from physicians who provide services to the public system (they might still be in the private sector, not public sector employees) might be easier than inducing the higher effort, and hence service quality, characteristic of private insurance companies who, in equilibrium, serve the non-poor. One might expect then to see relatively flat incentive schemes for physicians serving the public insurance system, and steeper compensation schedules employed by private insurers.

Another reason that public and private insurance systems may provide different incentive schemes to physicians is that physicians may represent a heterogeneous group. If they differ in their aversion to risk, in their ethical priorities, or in their job satisfaction, it may be optimal to offer one kind of compensation contract to one group of physicians, and another kind to a second group.

Once again, however, these arguments suggest reasons that we might see different ways of paying physicians emerge in the delivery of health insurance, but they are not necessarily the outcome of a public-private mix of insurance provision. For example, it is easy to imagine a purely private market becoming vertically segmented, with some plans targeting low-demand consumers (i.e., the relatively poor), and others targeting the non-poor, with a corresponding range of provider payment mechanisms.

### 3.2 Purchaser-provider split versus vertical integration

The previous section examined issues of how to pay physicians, under the assumption that such payment mechanisms were implemented through explicit contractual agreements. Instead of writing a detailed contractual agreement between insurer and physician, the two parties might decide instead to integrate into a single organization, and to rely on bargaining protocols to determine the allocation of rents. Traditionally, in many countries in Latin America and elsewhere, public insurance systems have been vertically integrated to a significant degree. Recent reforms however have focused on separating the functions of insurance and provision, through the so-called purchaser-provider split.
These reforms effectively require that explicit contractual arrangements govern relationships between insurers and providers.

On the other hand traditionally, private insurance was of the fee-for-service type (i.e., indemnity plans), whereby a physician would send a bill to the insurer for covered services. This is one kind — indeed a common one — of explicit contract. Over time, however, private insurance companies have moved towards a more integrated organizational structure, bringing physicians in-house, or at least adopting long-term contractual relationships with them. This apparent anomaly between the evolution of the organization of public and private systems can be understood in a number of ways.

**Soft budget constraints**

First, both institutional developments might represent attempts to provide physicians with stronger incentives, within constraints that differ between the public and private sectors. For example, it might be difficult for a government bureaucracy to commit to funding a public sector employee prospectively, thereby making him the residual claimant. Thus, if future renegotiation in the event of high costs makes it difficult to credibly threaten punishment, endogenous soft budget constraints limit incentives. Conversely, if political pressure is likely to force governments to expropriate profits (or to renegotiate future contracts) in the event of low costs, incentives will again be dulled. By contracting explicitly with the private sector — i.e., by adopting a purchaser-provider split — a public payer may be able to harden what would otherwise be an endogenously soft budget constraint.

On the other hand, the easiest way for a private payer to provide a physician with incentives might be to make him a shareholder in the insurance firm. (Of course, there are obvious free-rider problems when physicians are paid on the basis of group profits and not individual contributions thereto.)

**Common agency in public institutions**

---

4 It is important not to trivialize the nature of private insurance organizational structures. In reality a wide range of organizational forms exists, ranging from for-profit arms with shareholder doctors to non-profit HMOs with salaried providers.
A second way to understand the opposing directions of reforms in the public and private sectors is to look more closely at the nature of the purchaser-provider split arrangements. In particular, instead of contracting directly with individual physicians (as we assumed in the preceding paragraphs), a public insurance system might sign contracts with groups of physicians — indeed often with managed care organizations. This suggests that the function that is being contracted out from the public system is the management of physician services. Having a formal arm’s length contract between the public sector and the manager of physician services might be an effective way of improving the incentives of such a manager. For example, building on the work of Williamson (1985) and Hölmstrom and Milgrom (1990), Dixit (1997) has shown that when an agent (here the manager) reports to multiple non-cooperative principals with heterogeneous objectives, a negative externality exists amongst the principals that results in low-powered incentives for the manager. The implicit side contracts that may easily exist in a political environment support such a view of the constraints on incentives of public sector managers (but less so for individual physicians). Requiring an explicit contract could facilitate the cooperation of the principals, removing the externality, and leading to higher-powered incentives for the manager.

Explicit arm’s length contractual arrangements might also be a way of limiting the scope of a manager’s activities. The usefulness of limiting objectives of public sector decision-makers has been suggested by Tirole (1994) and formally elaborated in a model of career concerns by Dewatripont, Jewitt, and Tirole (1999). Providing a manager with a well-defined “mission” again makes it easier to induce effort. In Dixit’s analysis the narrowing of a manager’s objectives is effected by reducing the number of competing principals to whom the manager answers.

---

5 The model relies on the manager taking a number of actions or efforts. Principals preferences over the outcomes of different actions differ. Each principal will then propose a contract that offers the agent more insurance by making payment negatively correlated with performance in dimensions that the principal does not care for. All principals act the same, resulting in a dilution of incentives.

6 Contracting out to a private manager may not be necessary. Corporatization — an intermediate step — may be sufficient.
Contractual incompleteness

An alternative literature examines issues of contracting out versus internal provision — i.e., vertical integration — starting from the presumption that contracts are necessarily incomplete. Even if choices are observable by both parties to a contract, if they are not verifiable and if the contract is consequently unenforceable, then institutional arrangements can have substantive effects on incentives. In particular, ownership of productive assets can matter when explicit contracts are unavailable — private contractors are those who own the assets they use to produce services (e.g., hospitals), while public servants do not have the implied control rights over asset use.

Hart, Shleifer, and Vishny (1997) present a model of service provision when quality and cost are non-contractible. By definition, public sector employees cannot retain ownership of any quality innovations they generate, while private sector providers can. Incentives for quality innovations are thus greater in the private sector. On the other hand, incentives for cost reduction are also greater in the private sector, but cost control is associated with lower quality. Thus the private sector will always (in the model) produce at lower cost, but could produce higher or lower quality. When one of the ways of reducing costs is to actively select easy-to-treat patients ahead of expensive cases, the social aspects of quality can be severely effected by incentives to control costs.

One might be willing to argue that innovations in medical care are very important, while those in insurance administration are less so. This argument favors private provision of physician services, under the condition that active selection could be controlled adequately. However, in countries with large sections of the population uninsured against health needs responsive to standard treatments, innovation in insurance delivery may have high social payoffs, in which case public provision may then be favored.

If cost is non-contractible, then procurement contracts like those studied by La ont and Tirole (1993) are not feasible, and a fixed price contract must be used.
3.3 Competition

Competition within the public sector, in the private sector, and between the two, can provide incentives for quality provision and cost-reducing effort on the part of medical care practitioners. The important feature of competition is the discipline it allows consumers to place on service providers.

**Competition within government: quasi-markets**

Competition among suppliers should not necessarily be identified with private supply. Indeed, the UK government has attempted to induce competition among public providers by developing the so-called quasi-market (Le Grand (1991)). Even when consumers do not face financial incentives to choose wisely between suppliers, they might still induce effort and quality provision if their choices affect the payoffs to providers. Thus Halonen and Propper (1999) model the impact of competition on quality when providers are paid by a public sector payer on behalf of consumers who are free to choose their supplier. The essential feature of their model is that when providers’ objectives are not coincident with consumers’ (on average), allowing consumer choice can help to realign providers’ incentives. The benefits of competition are of course limited by the elasticity of demand.

**Private sector competition**

Competition can improve incentives, but can also have negative effects. The discussion above centered on public sector providers’ incentives to attract patients through quality improvements. An important aspect of quality is that, while it is reasonable to assume that all individuals value more quality than less, the willingness to pay for quality varies with health needs. In general, a hospital is likely to face a higher elasticity of demand with respect to quality from patients with high needs than from patients with low needs, assuming an alternative source of supply exists. Competition can then result in a kind of “race to the bottom,” in which all hospitals try to deter high cost patients from seeking treatment by under-providing quality. Such negative effects of competition would not arise if hospitals were able to charge sufficiently higher amounts for treating expensive patients. Similar forces are likely to be at work in private insurance markets.
In some private health insurance markets consumers are required to commit ex ante to limit their ex post choices, thus weakening the competitive pressures they can exert on physicians. This occurs under various managed care arrangements (HMOs, PPOs, etc.) These restrictions reduce ex post demand elasticities, allowing either prices to be higher, or more likely, quality to be lower.\(^8\)

**Public-private competition**

A common argument in favor of large purchasing groups is the monopsony power they can wield in negotiating supply contracts. However, as Propper and Green (1999) have recently pointed out, there is no particular merit in such actions — market power is inefficient be it on the supply side or the demand side. They suggest that under such arrangements staff will either be of poor quality (good staff will be driven from the market by low wages), or employment contracts will permit outside earnings with little or no monitoring, weakening incentives for performance of primary job tasks.\(^9\) *Introducing competition from the private sector* may thus lead to higher public sector wages and costs, but with a net welfare gain. The lesson from this analysis is simply that focusing on budgetary impacts alone is not sufficient to determine welfare effects, especially if rationing costs (i.e., waiting times, search costs) are included (Danzon (1992)).

Alternatively, *introducing public provision into a private market* may be beneficial if it serves the purpose of making a minimum quality standard credible. For example, Ronnen (1991) has shown in a model of vertical product differentiation that a minimum quality standard can raise the quality provided by all market participants (even those who would have met the standard in the absence of regulation), and lower equilibrium hedonic prices (i.e., prices adjusted for quality). An appropriately chosen standard, by restricting product differentiation and intensifying ex post price competition, makes all consumers better off. Such a standard could in principle be imposed by government, without recourse to public provision. If quality is difficult to monitor however, and if lapses are costly to penalize, public provision of the standard quality at a minimal price could act

\(^8\) For a review of models of strategic competition with horizontal product differentiation, see Tirole (1989).

\(^9\) Such arrangements are often found in academic institutions, particularly ones funded from public sources.
as a substitute for direct monitoring. In equilibrium, private sector quality would be higher than that in the public sector, and any increase in prices would not be enough to outweigh the positive effects of higher quality.

The idea that public and private sector quality can act as strategic complements should be viewed with some caution. Standard equilibrium analysis (Hammer (1997)) suggests the need to anticipate potential crowding out of private sector supply by public sector provision. In the case of non-differentiated goods this is indeed appropriate, and one expects that if the publicly provided quality level was too high then private supply would dry up.

4 Institutional evolution in Latin America — some examples

Over the past twenty years several Latin American countries have embarked upon wide-ranging reforms of their health insurance and delivery systems. Generally, among the higher-income countries there has been a move toward extending explicit insurance coverage to those outside the formal labor market. At the same time, these countries have examined the ways in which insurance and health care have been delivered, and instituted reforms that are meant to improve the allocative and production efficiency in the sector. They have tended to move away from integrated public provision of insurance and care, in the direction of more decentralized provision that relies, to varying extents, on private sector involvement. Major health insurance reform, like health care itself, appears to be a luxury good, and the poorer countries in the region have focused on more basic challenges in terms of primary care delivery. Below I will review the experiences of four countries that have adopted significant reforms: Colombia, Argentina, Brazil, and Chile.\(^{10}\)

\(^{10}\)For a more complete survey of health systems in Latin America see Bengoa et al. (1998).
4.1 Colombia

The health sector reforms initiated in the early 1990s under the direction of Minister of Health Londoño represent possibly the most ambitious policy interventions undertaken since the implementation of large-scale social security mechanisms of the 1950s and 1960s, both in Latin America, and outside the region. Before the reforms, Colombia’s public health delivery system was centralized, budget-financed, and poorly organized and consisted of two uncoordinated bodies: (1) the social security institutions (SSIs), that provided subsidized health services, and hence insurance, to formal sector workers; and (2) the Ministry of Health, that provided public health inputs, and subsidized hospital care for those who could not, or would not, use the SSI system. This sub-section will not present a full history of the “pre-existing condition” of the Colombian system or the reforms, but will examine how the reforms conform with some of the general organizational structures discussed in sections 2 and 3.

Consistent with the discussion in the first section of this paper, the general goal of the Colombian reforms appears to have been to ensure a basic level of coverage for all individuals, that could be improved upon for those willing and able to pay more. The redistributive nature of the reforms is manifest in two aspects: first, the tax base used to finance the new system is at least proportional, if not somewhat progressive (it consists of a payroll tax plus general revenues); and second, since in the absence of the reforms many informal sector workers and their families were formally uninsured the provision of a basic level of insurance at minimal personal cost acts like a negative income tax.\footnote{The uninsured did receive health care services under the old system, mostly from public hospitals and clinics. An assumption often made is that formal insurance will lead individuals to consume necessary care earlier, to have access to better quality care, and to face lower personal costs.}

There has indeed been a marked increase in formal coverage of the population, particularly among lower-income groups. Figure 3 (composed from Cuadro No. 7 of Uribe et al. (1999)) shows coverage rates by income quintile in 1993 and 1997. Overall, the proportion of individuals with insurance doubled in this period from 23.7 percent to 57.2 percent, with the largest proportionate gains among
the poor.

At the same time, the reforms attempt to improve incentives for productive and allocative efficiency, by encouraging alternative provider payment systems and allowing consumer choice, respectively. Integration of a market mechanism for delivery with redistributive objectives presents policy-makers with complex issues of design. Formally, there are three main groups of actors: consumers, purchasers, and providers. A fourth agent is the government, that, while retaining active roles in policy-making, regulation, and public health initiatives, acts as a budget balancer (in that it allows payments made by any one of the other three not to equal those received by the recipient).

**An implicit two-tiered voucher scheme**

Insurance coverage is allocated to consumers through two regimes with separate but related fi-
nancing arrangements. I will discuss these below, but it turns out these schemes are approximately equivalent to a simple two-level voucher system, although no formal voucher scheme exists. Effectively, members of one group of families (those with workers in the formal sector) receive a voucher for insurance that covers a wide range of services at notionally high quality, while individuals in the complementary set (essentially the poor) receive a voucher for a less generous package of insurance. Members in the first group are said to be in the “contributory regimen,” and those in the second are referred to as participating in the “subsidized regimen.”

On the financing side, participants in the contributory regimen pay a 12 percent payroll tax to help finance health care. Although this tax is earmarked for health services provision, because its payment is non-discretionary and the value of the implicit voucher received upon payment is unrelated to the amount of tax paid, we can, without loss of generality, consider the tax to be part of general revenue.\textsuperscript{12} That is, the distortionary effects of financing should not be considered to be offset by the fact that individuals receive something — insurance — in return for paying the tax.

Participants in the subsidized regime also contribute financial resources to support health insurance costs. These contributions are means tested (they may be zero in many cases) and not directly related to the cost of the coverage they receive. Since members of the subsidized regimen are required to purchase insurance, the contribution acts as part of the general tax base in terms of its distortionary effects. Thus despite the nominal link between financing and coverage at the individual level, it is appropriate to treat financing and demand as independent.

The next issue to discuss is where and how the implicit vouchers can be “spent”. Participants in the contributory regimen can use their vouchers to buy insurance from so-called EPSs (\textit{Empresas Promotores de Salud}), which are essentially private sector insurance companies. The EPS can cash in the voucher with the government (via the FOSYGA — \textit{Fondo de Solidaridad y Garantía}) for a fixed amount, known as the UPC, which is adjusted for some of the risk attributes of the consumer.

\textsuperscript{12}There are of course questions of optimal taxation to consider — is a wage tax more or less efficient than another revenue source, for example? I will not get into this here, however.
Participants in the subsidized regimen can use their implicit vouchers to purchase (less generous) insurance either from EPSs, or from ESSs (Empresas Solidarias de Salud). ESSs are medical care purchasing organizations that subnational governments must set up to facilitate the coverage of the self-employed and non-salaried workers who may not be able to access EPSs (say due to remoteness).

Because the implicit vouchers have fixed monetary values there is little effective price competition amongst EPSs and ESSs. They are allowed to charge a price lower than the UPC, but none of the price reduction accrues to consumers, so such a policy has little if any effect on demand. Instead, the EPSs compete on the basis of the level of insurance they provide, as determined by copayment rates, quality of service, and extra services offered. Thus, the standard packages of services defined for participants in each regimen act as basic plans, which can be supplemented by insurers to attract clients.\textsuperscript{13} This kind of arrangement encourages efficient provision of insurance, since the insurer gets to retain any efficiency gains.

However, there are two potential problems with using consumer demand to discipline insurance providers. First, if demand is not elastic in response to quality changes (i.e., if individuals cannot easily judge quality differences, and if they do not have easy access to alternative providers), incentives to control costs might outweigh incentives to improve or maintain quality. Given these determinants of demand, it is arguable that demand by participants in the subsidized regimen will be less quality elastic than that of participants in the contributory regimen, so that the quality of insurance provided to the poor will be reduced further.

Second, quality competition can naturally induce active selection efforts on the part of EPS/ESS managers, who might design policies to attract inexpensive clients, perhaps by bundling extra services, for which an additional premium can be charged, with elements of the standard package. If the standard package is provided at low quality, but additional services are provided at high quality, then it may be possible for firms to attract relatively profitable clients. For example, La Forgia

\textsuperscript{13}The standard package in the contributory regimen is referred to as the POS (Plan Obligatorio de Salud), and that of the subsidized regimen is the POS-S.
(1998, p. 257) reports concern over “the practice of EPSs to integrate the POS [standard package] with complementary plans, thereby undermining competition for a homogeneous service plan.” Both of these concerns can in principle be addressed by regulating more closely the quality of insurance provided, but there are obvious limits to the efficacy of such policies.

An intriguing aspect of the Colombian system is that two basic levels of insurance are mandated for separate sections of the population. Mitigation of adverse selection problems and public provision (or at least mandatory consumption) of a private good to facilitate redistribution as discussed in the first section of the paper each imposed a single insurance package on all individuals. The resolution of this apparent anomaly is that although a base level of insurance might be useful in redistributing income from rich to poor, within the group of individuals who would purchase supplementary insurance under such a system significant heterogeneity of risks will likely persist. In order to facilitate such voluntary supplementation while avoiding the potential adverse selection problem within this (wealthier) group, another specific insurance contract is imposed.\textsuperscript{14}

\textbf{The supply side}

The efficiency of provision is determined by how providers are paid, and how they are organized. This section briefly reviews the financial arrangements that govern incentives and the degree of integration of insurance and provision.

\textit{Demand versus supply subsidies}

On the supply side, EPSs and ESSs contract with hospital and physician groups, including private sector IPSs (Instituciones Prestadoras de Servicios), and formerly public sector, but now autonomous, ESEs (Empresas Sociales del Estado). The financial aspects of such contracts are less precisely regulated than are the demand side transactions between consumers and EPSs/ESSs,

\textsuperscript{14} A simpler explanation for the use of two levels of mandatory insurance is that it is too expensive for the government to enhance the POS for the poor. Intentions are to phase the POS-S up to the POS over time. However, if it were only a matter of money, the question would then arise as to why a uniform coverage plan, below the POS but above the POS-S, was not initially introduced for all individuals, and then phased up over time, with the same time profile of budgetary expenditures. Presumably the answer to this question lies in the political realities of reform.
although the law attempts to encourage innovative payment methods to encourage provider effort and efficiency (e.g., capitation, DRG-based payments).

One aspect of the reforms that has received some considerable attention is the fact that medical care providers are paid on the basis of demand. That is, insurers contract with providers to supply services for the covered clients, based on expected (in the case of capitated payments) or realized (in the case of fee-for-service and DRG-based contracts) use. In the past, when the government paid for services, payments were nearly always purely prospective, deriving from budgetary allocations to hospitals, provincial health ministries, etc. The assumption of many commentators appears to be that such a transition is undeniably beneficial.

However, there are two dimensions along which we might expect the alternative regimes to differ. First, in terms of the use of a fixed revenue base, demand-based financing would appear to offer the prospect of a more efficient allocation. On the other hand, due to the well-documented price elasticity of demand for health care, prospective budgetary allocations have the advantage of imposing a cap on aggregate spending. Indeed, before the reforms Colombia exhibited a relatively low level of health expenditure amounting to around 4 percent of GDP in 1990, but recently this has exploded to nearly 11 percent of GDP in 1998. There is some suggestion that this expenditure increase has been caused by the introduction of demand subsidies without the removal of supply subsidies (so hospitals receive budgetary funds and can additionally charge EPSs and ESSs for services provided), so that the welfare implications are unclear. Nonetheless, to achieve allocative efficiency both within the health sector and across sectors, provider payment mechanisms must be carefully designed, incorporating both demand side responsiveness to needs, and constraints on

\[15\] The welfare effects derive from two sources: the extent to which the additional expenditure finances inefficient production versus private consumption (of medical personnel or hospital owners); and the shadow cost of public funds. That is, if the additional expenditures are not matched by increased service levels (which could be of little value), they represent pure transfers with little efficiency effect. Except, that is, if the shadow cost of public funds is not considered. When the distortionary costs of raising the additional revenues are incorporated, the welfare impact of the increased spending, no matter what it finances, can be large and negative.

25
over-use.

**Purchaser-provider split**

The formal separation of purchaser and provider is mixed in the Colombian system. On the one hand, under the proposed reforms (when fully implemented) the government will have little direct role in the provision of either health insurance or health care. The FOSYGA acts as a clearing house for taxes paid by individuals and transfers made to EPSs and ESSs. Apart from monitoring the quality of the goods (i.e., insurance packages and medical services) provided, the public sector does not actively perform any purchasing role, this function being delegated to consumers.

On the other hand, the degree of integration of the delivery of insurance and medical care varies widely. Some purchasers (EPSs and ESSs) contract at arm’s length with provider networks (IPSs and ESEs), while others effectively own such networks, along the lines of HMOs. Even if an EPS owns or has close contacts with a particular provider network, it is required by law to offer the services of at least one other IPS to consumers. On the face of it, since consumers have free choice of EPS, requiring that each EPS offers services through at least two IPSs seems redundant. One possible explanation for this regulation is that it aims to increase ex post quality competition (among providers), under the assumption that it is costly for individuals to changes EPSs, but less so for them to switch between IPSs offered by any given insurer.

### 4.2 Argentina\(^\text{16}\)

Compulsory tax-financed health insurance has been delivered in Argentina by a plethora of for-profit and non-profit institutions in conjunction with a predominantly private supply of physician and hospital services for many years. On the face of it, this might appear to be a model with high potential, dealing with insurance market failures (e.g., adverse selection) through compulsory participation, equity concerns through progressive taxation, and efficiency of delivery through the power of decentralized provision of insurance and services. However, a number of factors have meant

\[^{16}\text{Much of this sub-section relies on World Bank (1997, 1999).}\]
such potential has not been realized, and have led to major reforms of the sector in the mid- and late-1990s.

**Descriptive analysis**

Argentina’s formal health insurance system in the late 1980s was well-developed. All employees were obliged to be insured by the so-called *obra social* that covered the sector in which they worked. In essence, the *obras sociales* were, and remain, non-profit insurance companies owned by the relevant labor union. There are about 360 such *obras*, covering about 10 million individuals and their families. In addition, each of the 24 provinces of the country operates an *obra provincial*, covering about 5 million public sector employees and their dependents. *Obras* are funded on the basis of (non-discretionary) payroll taxes.

Retired workers and pensioners — about 4 million individuals — received health insurance coverage through the Integrated Program of Medical Care (PAMI, *Programa de Asistencia Médica Integral*), operated by the National Social Service Institute for Retirees and Pensioners (INSSJP, *Instituto Nacional de Servicios Sociales para Jubilados y Pensionados*). These services were funded by payroll taxes and taxes on pension benefits. The *obras* and PAMI combined covered about 61 percent of the population. Another 2 million people were covered by private for-profit, pre-paid insurance plans (*pre-pagas*), and a further 1 million received insurance through small *mutuales*, of which there were around one thousand. The distribution of coverage across types of insurance is shown in Figure 4. The 25 percent of the population who lacked formal insurance would typically seek care through public hospitals operated by provincial and municipal governments.

Most individuals with formal insurance sought the care of private sector providers, although *obras* provided on average 10 percent of services in-house. Some *obras* in fact contracted out the insurance function as well, paying *Empresas Administradora de Prestaciones* (EAPs) a capitation payment per member: the EAPs then contracted with provider networks for service provision. Traditionally providers were compensated on a fee-for-service basis, although diagnosis-based payment and capitation have become more common.
Argentina Health Insurance: Types of coverage

Figure 4:
Individuals covered by PAMI were assigned a family doctor, who was supposed to act as a gatekeeper and was an INSSJP employee. Other providers (specialists, hospitals, diagnosticians, etc.) are paid with a mix of capitation payments and fee-for-service contracts.

Recognizing the wide variation in both wages and health risks by occupation and hence the wide variation in per capita contributions across obras, the Administracion Nacional del Seguro de la Salud (ANSSAL) in theory provided redistribution both from rich to poor obras, and from low-risk to high-risk obras. ANSSAL also provided a regulatory structure within which the obras operated.

Outcomes

The main problems identified with the provision of health insurance were consumer dissatisfaction with poor quality (i.e., allocative inefficiency), financial instability (deriving from unsustainable production inefficiency), and inequity. We address each of these in turn, and describe the associated reforms.

Allocative inefficiency

Discipline on providers comes either from the demand side through consumer choice, or from effective regulation. Under the old system, formal sector workers were not permitted to exercise choice on any dimension. They could not choose how much to pay for insurance, because the payroll tax was non-discretionary (and opting out was not permitted), or from whom to purchase it. INSSJP beneficiaries were not permitted to choose their family doctor. While there are well-understood limits to the efficacy of consumer demand responsiveness in health care and insurance markets, the abolition thereof is unlikely by itself to improve matters.

Alternatively, the regulatory apparatus of ANSSAL and INSSJP was not sufficient to ensure quality coverage in the absence of consumer choice. Some regulatory failures were obvious, for example the ability of obras to engage in activities not related to health insurance (like recreation facilities). Others derived from the political use to which executive positions in the regulatory institutions were put and the scope for corruption that emerged.

Financial instability
Financial instability derived both from the revenue and cost sides. Total wage taxes (of which health insurance contributions constituted about 20 percent) amounting to 56 percent of the gross wage were deemed to be excessive and an impediment to macroeconomic performance and international competitiveness and were reduced during the early 1990s, cutting into the health insurance tax base. The 1995 financial crisis and subsequent economic slump did not help matters, to say the least. Of course, as a policy instrument, such tax cuts should not be seen as a cause of production inefficiency, but they did contribute to mounting debts amongst the obras. On the cost side, weak management, excess staffing levels, and rising medical care costs contributed more fundamentally to the inefficiency of resource use.

The power of the trade unions, as owners of the obras, was important in this respect. World Bank figures show average beneficiary-staff ratios of about 130, with some obras reporting fewer than 10 beneficiaries per employee. Despite the high level of unemployment, and the associated low shadow wage, the enormous distortionary costs associated with such labor intensive production would appear to outweigh the positive welfare effects through their impact on the unemployment rate.

Weak management of the obras may indeed be a misrepresentation. In fact, managers exercised significant control of resource uses and policy choices, usually in favor of the controlling union and not necessarily in the long-term interests of the workers. Without any incentive to control costs — both because there were no shareholders, and because the government imposed a soft budget constraint — otherwise productive effort was diverted to political maneuvering and rent-seeking.

Even with little consumer-imposed discipline, one might have expected the obras to aspire to productive, if not allocative, efficiency. As non-profit organizations any surpluses generated would have to be distributed in kind (this may explain the involvement of the obras in non-medical activities), but this should not have dampened cost-consciousness unduly. Instead, costs escalated while service quality remained stagnant at best. In fact, this situation arose because, far from being residual claimants vis-a-vis the net revenue streams generated from insurance provision, obras faced
very soft budget constraints due to easy access to government credit. Since individuals had no possibility of switching obras, even in the event of financial collapse, it was difficult for the government to commit to allowing them to go under. Indeed it was far more politically palatable (and in the short run, possibly more efficient) to extend credits and permit continuation of coverage with the existing obra than to close down an insurer and assume responsibility for coverage of its members. Thus the lack of formal and credible bankruptcy procedures meant the threat of shut-down was an ineffective incentive mechanism.

Redistribution

About 90 percent of payroll tax contributions were retained by the obra in which an individual worker was enrolled and the remaining 10 percent of revenues passed on to ANSSAL for redistribution. This redistribution was meant to be both from rich to poor obras, and from low-risk to high-risk obras. In principle, such redistribution would allow the different obras to provide broadly similar levels of coverage.

However, in 1994 per-beneficiary own revenues varied from $5 per month for the poorest obra to $80 per month for the richest. Only a token impression could hope to be made in such gaping inequality with the resources available to ANSSAL from the payroll tax. In practice, things were worse, with per-beneficiary transfers being only slightly negatively correlated with own income (World Bank (1997)), suggesting only a marginal, if any, impact on inequality of health insurance coverage among workers.

From this discussion it can be inferred that obra revenues took the form of a more or less fixed payment per person plus a proportion of the average wage in the sector. Thus there was some redistribution within each sector, but not across sectors. This can be compared with the payment received by Colombia’s insurance providers — the EPSs — whose per capita revenue is a fixed amount (based on the UPC), with no proportional component. At the other extreme, Chile’s ISAPREs (see below) are funded more or less in proportion to each individual’s wage, with no fixed component.

The other redistributive function of the ANSSAL was meant to allow obras with different under-
lying costs, associated with different risk profiles, to provide similar service levels. Since membership in a specific obra was mandatory, and obras could not reject participants, risk profile differentials contributed only to inequality, and did not have any direct efficiency impacts associated with active risk selection etc.\textsuperscript{17} To the extent that individual risk is correlated with personal income, the offsetting risk adjustments could thus in principle have been incorporated directly into the income-based redistributive mechanism. Instead, transfers were made not on the basis of ex ante risk, but on the basis of ex post realized costs, in particular for certain procedures that were considered high-cost (although some, like prostheses, were cheap but still included).

**Current and future reforms**

Over the long term, insurance reform is aimed at attaining universal coverage through competitive provision of at least a minimum level of insurance. The existing fragmented structure — consisting of a public system for pensioners and retired individuals, a non-competitive but decentralized system of obligatory insurance for formal sector workers, an unregulated competitive system for others able to pay, and a public hospital system providing insurance through the provision of low quality services — is to be unified across consumers. That is, the characteristics of insurance providers and the environment in which they operate, are to be independent of the identities of the individuals they cover. This represents a rational separation of the organization of production (determined by internal efficiency considerations), and the allocation of consumption (determined by allocative efficiency and equity concerns). The major incentive instruments are to be consumer choice and, where this is ineffective or imperfect, regulation, both of quality and financial soundness.

\textsuperscript{17}Risk adjusters, even with mandatory insurance and no choice of insurer, could have efficiency effects in the labor market. For example, consider a situation in which individuals are taxed a fixed amount, and must obtain coverage from their employers. High-risk individuals will prefer more insurance and lower wages to low risks, and the two may sort themselves into different occupations, especially if cash wages within an occupation are inflexible. Risk adjusters can remove the link between wages and the value of insurance, and allow a more efficient allocation of labor across sectors. When individuals pay a fraction of their wage for insurance (and not a fixed amount) this story has to change a bit.
Competition among obras is being introduced, and by mid-1999 most people were aware that they had a choice of insurer. Consolidation in the industry is clearly required, and has been partially achieved through mergers of obras. Insurers are required to offer a standard health benefits package (the PMO — Programa Medico Obligatorio). As well as representing a minimum level of insurance that may facilitate redistribution in general, the PMO has enabled consumers to make relatively easy comparisons between obras.

Competition between obras and pre-pagas is so far limited. Individuals who are eligible for coverage by an obra are not permitted to obtain formal insurance from a pre-paga. However, some schemes have developed whereby an obra will subcontract with a pre-paga, thus allowing an individual to formally remain in the obra sector while effectively receiving insurance from outside.

Through the solidarity fund, obras with low incomes are assured of a minimum level of funding. Of the individual contributions of 8 percent of wages, the obra receives 90 percent. The additional 10 percent is used to finance transfers to poor obras in order to bring them up to a minimum of $40 per family per month. Thus the net price paid by individuals is still strongly positively correlated with their wage, so it is likely that some equilibrium vertical segmentation by quality will result. Part of the reform process is to change the $40 per family minimum to a risk-adjusted (by age and sex) per person amount. This would increase the redistributive role the solidarity fund (assuming noone got less than $40 under the new scheme), reduce the proportion of the wage tax directly retained by the obras, and reduce the potential for vertical segmentation of the market.

4.3 Brazil

Health insurance in Brazil is something of a mix between a nominally comprehensive public system and a large and active private system. Unlike Argentina, Chile, and Colombia that have tried in var-

---

18 The number of registered obras fell from 360 in 1997 to 294 in 1999.
19 For white collar workers the retained portion is 85 percent.
20 The 10 percent also funds transfers for high complexity and long-term treatments, as well as some other transfers to obras in special situations.
21 Much of this section is based on information reported in World Bank (1994) and Lewis and Medici (1998).
ious ways to implement (at least partially) privately provided mandatory insurance coverage, Brazil has opted to provide universal public insurance (financed from general revenues) while encouraging use of the private sector as an alternative. Consumers certainly have the freedom to choose among private plans, and are able to choose whether or not to use the government system. The big difference is that when opting out of the public system, individuals do not take their financing with them. As a result consumer choice provides little incentive for the public system to maintain quality. In reality the public system acts as a floor, available to all but used primarily by the lower classes.

The health sector reforms have concentrated on the organization of public service delivery, particularly through decentralization, and the payment of providers by the public system. Until recently, relatively less attention has been paid to the demand side, either with respect to using demand as an efficiency-enhancing mechanism by exposing the public sector to competition, or with respect to consumer protection through regulation of private insurance.

The public insurance system

Until 1993 public insurance was implemented through INAMPS (Instituto Nacional de Assistência Médica e Previdência Social), the medical arm of social security. INAMPS originally covered contributors to social security, i.e., formal sector workers, but under the new constitution of 1988 and subsequent legislation coverage was nominally extended to the whole population. Having a universal right to health care, as enshrined in the constitution however, has not helped those in the northern states who continue to lack adequate medical care facilities. Public health insurance expenditures are financed through earmarked taxes (e.g., a tax on the gross income of enterprises — Contribuição Financeira para o Setor Social, or COFINS, and a 0.3 percent financial transactions tax — Contribuição Social sobre a Movimentação Financeira) and general tax revenues of the federal, state, and municipal governments.

A comprehensive reform of the health system (Reforma Sanitária) was instigated in the early 1980s. The first phase of this, in 1984, was the integrated health actions (AIS) reform, which shifted some supply to under-utilized public hospitals, and coordinated the functions of INAMPS
with the Ministry of Health. Later, in 1987-8, a further reform (SUDS — Sistemas Unificados e Descentralizados de Saúde, or Unified and Decentralized Health Systems) led to a transfer of INAMPS staff and facilities to state health secretariats, with the central agency acting solely as a funding conduit (and being renamed the INSS — Instituto Nacional de Segurança Social). This decentralization reform was partially reversed in the third phase in 1990 under the SUS reform (Sistema Único de Saúde — Single Health System), with INAMPS reverting to a more centralized structure, although state governments were given authority to write contracts with private providers. In 1993 INAMPS was abolished and integrated into the Ministry of Health under the umbrella of the SAS, Secretaria de Ações de Saúde. The main function of the SAS is to transfer funds to state health secretariats.\footnote{Some of this transfer is made prospectively on a per capita basis for the purposes of primary care.}

INAMPS implemented most of its insurance coverage through contracting with private sector providers, first on a fee-for-service basis, but later using a U.S. Medicare-type prospective payment system (PPS). There are currently two PPSs in operation, for outpatient and inpatient services respectively.\footnote{I think the PPS for out-patients is actually a per capita transfer from the central government to the states, which is then used to eance care. But I’m not sure what the relationship is between usage and payment — i.e., if that is truely prospective or not.} There have been signi cant problems with data management under these schemes — indeed, the cost control attributes of prospective payment have probably been ine ective because of a lack of monitoring and evaluation by the public payer.\footnote{Let me explain: if diagnoses are inelastic and veriable, then PPSs provide good incentives for cost containment (although they may still overpay for cheap cases, see Chalkey and Malcomson (1999)). However, if the payer does not observe the validity of the diagnosis or subsequent resource use, providers can increase their revenues by over-diagnosing. In theory, the incentives for cost control remain, as the provider continues to be the residual claimant, and the social cost increases only due to the distortionary costs of raising the additional revenues to pay the higher charges. However if the higher revenues can only be consumed in certain ways (e.g., through the purchase of high-tech equipment) then production ine ciency adds to the social cost.}

The prospective payment of medical providers and hospitals on the basis of utilization suggests that the Brazilian government has chosen a limited version of contracting out. This reimbursement...
mechanism does not represent the contracting out of the management of individuals’ health care needs in any meaningful sense, as might be descriptive of capitation funding mechanisms. Nor does it represent the contracting out of insurance, which again could be implemented through a (risk-adjusted) capitation payment to an intermediary that would provide coverage for the individual of agreed services with associated copayments, etc. Thus while the private sector is heavily involved in the delivery of services financed by public insurance, the incentives of providers to focus on outcomes (i.e., health) as opposed to inputs, are relatively weak, with predictable impacts on quality and costs.

Private insurance

Private insurance covered about 25 percent of the population in the mid-1990s. The expansion in private coverage over the past 30 years has been brought about by a number of factors, including growing incomes (in the 1970s), a tax deduction of out-of-pocket expenditures and premiums, and the deteriorating quality of the public system.

Private insurance is provided through four alternative types of organizations. The largest, and historically most important, is the prepaid group practice (PGP), which is similar to the U.S. staff model HMO (i.e., physicians are salaried employees of the insurer), and had (in 1991) about 47 percent of the private insurance market. Medical cooperatives (like IPA — Individual Practice Association — models in the United States) restrict use to a set of preferred providers with whom contracts are written, and captured 28 percent of the market in 1991. Large employers (20 percent of the market) at times self-insure and offer company health plans, sometimes contracting out the administrative functions to financial intermediaries. Finally, only a very small proportion of those covered by private insurance (4 percent) enroll in indemnity plans (i.e., reimbursement insurance).

Regulation of the private insurance market was virtually non-existent until 1998. Indeed, in the early 1990s a request by the prepaid group practice trade association (ABRAMGE, whose members represent 45 percent of the PGP market) for public regulation was turned down. Thus exclusions and restrictions are common, financial soundness unchecked, and fraudulent practices are at least
perceived to be prevalent.\textsuperscript{25} In fact, the poor reputation of PGPs has been the main factor contributing to the relative growth of medical cooperatives and company health plans over the past 10 years, although PGPs still capture the largest portion of the market. Lately however, a number of regulatory initiatives have arisen. These include a regulatory framework created by the federal government, and the establishment of a Supplemental Health Department within the Ministry of Health and a Supplemental Health Board (consisting of representatives from private insurance companies, government, and consumer groups) to improve monitoring and control. An independent regulatory agency is currently being considered in conjunction with technical assistance from the Inter-American Development Bank, to cover issues of public information, solvency, accreditation of plans, quality, consumer rights, and risk management.

**Meeting efficiency and equity objectives**

It could be hoped that Brazil's universal public insurance system complemented by an active private insurance option, and its heavy reliance on private sector provision of medical care, could strike a balance between the efficiency and equity concerns outlined in the first section of this paper. However, despite the ambitious nature of reforms over the past 15 to 20 years, the system falls short of achieving this goal. On the one hand, to be meaningfully implemented, a universal base level of public insurance must be supported by medical care services that are in fact available to all. However the quality of and access to publicly financed services varies widely, suggesting that the public system is insufficient to meet the needs of the poor.\textsuperscript{26}

The efficiency attributes of private delivery on the other hand have been compromised by exactly

\textsuperscript{25}The public bad nature of a reputation for poor performance explains ABRAMGE's request to the authorities for external regulation. Effective self-regulation would have required the establishment of a reputation, that, given the behavior of some PGPs, was costly to achieve. Using an outsider to gain credibility would have been cheaper, but was not attractive to the government.

\textsuperscript{26}An alternative interpretation is that the public system is too generous for the non-poor, but given that the poor often pay relatively high prices to use the private system suggests that any base level of public insurance is essentially worthless to them.
those characteristics of health care and health insurance that suggest caution in relying too heavily
on the market to generate quality and cost savings. Unregulated, private insurance companies
have provided limited insurance (no lifetime coverage, excluded conditions, and other limitations on
coverage), and despite avoiding the fee-for-service paradigm have been unable to control costs.

4.4 Chile

Chile’s health system is one of the more closely and thoroughly studied in Latin America. The
country undertook wide-ranging and innovative reforms of its health and social security systems
starting in the early 1980s partly in response to failures of the bureaucratic centralized regimes
that had prevailed previously. Similar to the arrangements in Argentina, the reformed Chilean
system of compulsory health insurance couples private provision of insurance for some with public
insurance for others. Funding is primarily through a proportional payroll and pensions tax, although
there is no equivalent of Argentina’s redistribution fund. Colombia’s reforms have been more
ambitious, implementing more fully a redistributive allocation mechanism, while privatizing, or at
least corporatizing, insurance provision to a greater degree. The pattern of insurance coverage by
income and risk category that has resulted in Chile has been criticized by some commentators.
Nevertheless this pattern could be appropriate under certain assumptions about the redistributive
capacity of the tax and transfer system.

The mix of public and private insurance

Under the reformed system, formal sector workers and pensioners are required to contribute 7
percent of their incomes to finance health insurance. Each individual has the choice to allocate
his/her contributions to one of about 35 private insurance companies, known as ISAPREs, or to the
Fundo Nacional de Salud, FONASA (National Health Fund). The designated recipient of the funds
then provides insurance coverage for the individual and his/her dependents. Individuals who are

---

27 This sub-section relies heavily on Larraaga (1997) and Bitr n (1998).
28 There might be a reinsurance fund, but this is different.
unemployed or who work in the informal sector are automatically covered by the public insurance system.

The private insurance sector consists of about 35 ISAPREs,\textsuperscript{29} covered about 31 percent of contributors in 1995, and accounted for about half of insured medical care spending. ISAPREs can offer multiple policies, and are free to charge corresponding premiums. These premiums can vary on the basis of age, gender, and the number of insured, and on the quality and extent of insurance, for example as defined by the extent of coinsurance and copayments.\textsuperscript{30} Individuals are permitted to increase their contributions above the 7 percent of income in order to purchase a higher cost policy. In 1995 there were close to 9,000 policies offered by ISAPREs, reflecting a near continuum of vertical differentiation matching the distribution of wages and hence contributions.

One important characteristic of contracts offered by ISAPREs is that they often include stop-loss components, limiting the financial risk of the insurer to a certain amount (above which the individual receives no reimbursement). This feature is perverse. Normally we think of optimal insurance as covering catastrophic events nearly in full, but with demand-controlling cost-sharing in the form of deductibles and/or coinsurance at lower levels of expenditure. Indeed, the stop-loss feature, more than explicit restrictions on covered services, limits the extent of insurance actually offered, and means that the high-risk — especially the elderly — choose not to participate. Table 1 reports the shares of each age group enrolled in FONASA and the ISAPREs.

Of course, those with low incomes also tend to choose not to participate in the ISAPRE system because the premiums are unaffordable, and because the public system provides free insurance.\textsuperscript{31} Figure 5 confirms this pattern of demand, although it has changed over time. FONASA classifies

\textsuperscript{29}In 1997 there were 34 ISAPREs: 21 were open to any purchaser, and 13 were closed funds, open only to employees of specific firms.

\textsuperscript{30}ISAPREs cannot discontinue insurance, and can impose at most an 18-month waiting period on clients for pre-existing conditions.

\textsuperscript{31}In terms of the choice between private and public insurance, the latter is free even for contributors, since their contributions are non-discretionary. The premiums charged by private insurers are higher than 7 percent of income, but are of course discretionary.
<table>
<thead>
<tr>
<th>Age group</th>
<th>0-1</th>
<th>2-14</th>
<th>15-24</th>
<th>25-54</th>
<th>55-64</th>
<th>65+</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>FONASA</td>
<td>65.8</td>
<td>64.7</td>
<td>58.3</td>
<td>57.0</td>
<td>68.3</td>
<td>79.9</td>
<td>61.8</td>
</tr>
<tr>
<td>ISAPREs</td>
<td>24.9</td>
<td>23.9</td>
<td>22.4</td>
<td>26.7</td>
<td>16.5</td>
<td>6.9</td>
<td>23.1</td>
</tr>
<tr>
<td>Other</td>
<td>9.3</td>
<td>11.4</td>
<td>19.3</td>
<td>16.3</td>
<td>15.2</td>
<td>13.2</td>
<td>15.1</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Source: Uribe et al. (1999) Table 2.1

Table 1: Chilean health insurance: Distribution of coverage by age

beneficiaries into four income groups, labelled A (the poorest) through D (the richest). (It also has a fifth classification, E, for non-beneficiaries, i.e., those enrolled with an ISAPRE, or not formally registered with FONASA.) In 1994, half of FONASA’s beneficiaries came from the lowest income group, 25 percent from the next group, and around 12-13 from each of the richer two groups C and D. In 1996 this picture started to change, with individuals from the poorest group representing just one-third of FONASA enrollees, while the share of group D enrollees doubled to about 25 percent. The shares of groups B and C however did not conform with this convergent tendency.\(^{32}\)

In most dual public-private systems, the coverage offered by the public system is independent of income, although in practice some correlation is to be expected if public facilities in poorer areas are of lower quality. However, in Chile this is not the case. In fact, there is in principle a negative correlation between income and coverage for those choosing the public system because copayments of between 0 and 50 percent are charged on a means-tested basis for hospital services.\(^{33}\) Thus, holding risk constant, the value of public coverage falls with income. Since the value of endogenously priced private insurance increases with income, the income profile of coverage is U-shaped. This design

---

\(^{32}\) This discussion is incomplete as it has not been established that the income categories represent quartiles (i.e., with equal numbers of individuals across the population).

\(^{33}\) Larraaga (1997). This negative correlation may be offset by the fact that contributors to FONASA are eligible for services provided by private physicians, with a copayment, whereas the indigent are required to use public providers.
induces more individuals to opt for the private system, but reduces the value of coverage for those at the margin who stay in the public system.

The resulting pattern of coverage in which those with high risks and low incomes use the public system and others use the ISAPREs has been criticized as fragmentary. This is certainly the case, but it does implicitly assume that uniform coverage is optimal. While this is probably true in the first best (abstracting from differences in risk preferences and demand elasticities), in a second best world in which the government wishes to redistribute from the rich to the poor and from low risks to high risks, such a pattern may not be unreasonable.

What is perhaps questionable is the extent to which higher-income individuals are effectively required to purchase superfluous insurance, or at least insurance of low marginal value. A comparison with Colombia’s more focused system is useful in this respect. Under current practice, Colombian users of the contributory regimen (corresponding to Chilean ISAPREs) receive a more or less stan-
dard insurance package independent of income, implying an in-kind redistribution from high-income contributors to lower-income contributors, which results in two tiers of more or less uniform health insurance. In the ISAPRE system however, above the public level of insurance there is a continuum of insurance qualities, increasing with income. The Colombian arrangement therefore allows for some redistribution within the group of relatively higher-income individuals. Perhaps more importantly, since the average value of the implicit voucher used in the ISAPRE system is equal to the average contribution across all ISAPRE users, there is limited redistribution (through the health system at least) between groups in Chile. In contrast, because the cost of the standard insurance package under the Colombian contributory regimen need not (in fact, does not) equal the average contribution by users, there is scope for further inter-group redistribution.34

A final characteristic of insurance contracts generated under the ISAPRE system is that, since they are tied closely to wages, quality tends to follow the life cycle pattern of wages, which may be somewhat different from the time profile of an individual’s demand for insurance. Thus it is not just the case that some individuals with high lifetime earnings will be induced to purchase more insurance over their life spans than they would desire, but that the pattern of coverage may not match their pattern of needs. Even if capital markets work well, this problem will persist, unless insurance companies can write long-term contracts in which high contributions at ages of high earnings but relatively low risk are made in exchange for lower contributions later in life when earnings are lower and risks higher.

34 Of course, to make definitive statements about the extent of redistribution between groups we need to know the full structure not only of taxes, but other government expenditures. Focus on the earmarked taxes and the provision of insurance is useful for predicting the effects on distribution of marginal parameter changes. For example, an increase in the payroll tax rate in Colombia would increase the resources available for both the low quality and high quality insurance packages. Doing the same in Chile would necessarily increase the quality of the high quality packages, with probably little effect on that of the publicly provided services. On the other hand, Bitrén (1998) has found that within the FONASA-eranced public insurance, the incidence of net benefits is reasonably progressive, suggesting a degree of within-group redistribution through the public system.
The organization of publicly provided health care

About 70 percent of Chile’s population is covered by the public insurance system, which financed care delivered through mainly public hospitals and a mix of public and private ambulatory care. Medical services financed through the public budget are funded primarily from FONASA and the general budget (for the indigent), with additional resources deriving from copayments by public patients, payments from ISAPREs that used public facilities, and other sources. FONASA acts primarily as a decentralized financing agency, collecting contributions and distributing funds to providers through a network of 26 Health Services. Following a contraction in public health spending in the 1980s (reflecting in part, but not fully, the growth of ISAPREs during that period), real public health expenditures more than doubled between 1990 and 1996 (Larrañaga (1997)). Despite this increase, the perceived quality and adequacy of public services remained stagnant, leading to a recent debate about reform of the internal organization of the public system. Some have argued in favor of moving towards internal markets (see section 2) and introducing competitive pressures into the delivery system, while others — particularly some labor unions and parts of the medical profession — have resisted what they see as a move towards privatization.

Although a decentralized organizational infrastructure exists, in the form of the 26 Health Services, central bureaucratic control of some crucial decisions remains in the hands of the central authorities, including the Ministry of Finance. In particular, labor inputs and compensation are dictated from the center and funded directly from the budget. Some central control of manpower allocation is likely to be necessary in the health system, especially one in which providers are not necessarily profit maximizers and in which service provision has a redistributive role. However, allocations made centrally tend to be unresponsive to changing cost structures and needs, and provide little incentive for innovation.

One part of the financial apparatus that does provide incentives for cost-consciousness is the use of a prospective payment system for financing the use of drugs and material supplies.\(^{35}\) As is well

\(^{35}\)This is referred to as the PAD. Less common services are funded on a reimbursement basis from a global budget
known, prospective payment gives good incentives for cost control, although as discussed in section 2 it may induce undesirable actions by providers on other margins, the most common identified issue being quality. However, even in the absence of quality concerns, the cost containment incentives are only operative if providers can exercise choice over input use. When the major component of costs — labor — is not under the control of the hospital, the impact of incentives for cost control is reduced significantly. Indeed, it could lead to a bias in favor of admitting patients with conditions whose treatment is relatively labor intensive, since the shadow cost of labor to the hospital is effectively zero, at least in the short to medium run.\footnote{More specifically, it could lead to a bias in favor of conditions for which the elasticity of substitution between labor and non-labor inputs is high, in which case the return to a switch towards a more labor intensive production process is relatively high.}

As well as increasing the effectiveness of incentives for cost control by allowing providers (hospitals) to become residual claimants, a movement towards full prospective payment (that would fund wages as well as other inputs) could also generate a degree of competition similar to that envisioned in the quasi-market reforms of, for example, the United Kingdom. Whether this competition is socially beneficial depends, as usual, on the responsiveness of consumer demand to quality, both as determined by geographical constraints and limitations on information. When it is not, which is especially likely the case in poor and/or rural areas, direct monitoring and regulation of quality are necessary.

5 Concluding comments

Despite the well-known failures of insurance and health care markets associated with imperfect information, the primary motivation for large-scale public intervention in the sector has been equity. Most often this derives implicitly from a concern for good-specific equity, but it can also be justified under a mechanism known as PPP. When the global budget constraint binds, such services must be rationed in some fashion.
as part of a more general second-best redistributive mechanism. Within the context of public intervention in the pursuit of equity goals, it is reasonable to assert that the reforms of the last two decades have sought not so much to improve the efficiency of private markets, but to improve the efficiency of public provision, either through direct use, or mimicking, of such markets. This has sometimes been achieved by altering the focus and function of pre-existing institutions — e.g., the obras sociales in Argentina, or by encouraging the growth of new institutions, such as the ISAPREs in Chile.

Coupled with the reforms of the ways insurance and care are organized and delivered, countries have attempted to various degrees to extend formal coverage to previously marginalized groups, and to finance this extension fairly. Colombia can be seen as instituting an implicit two-tiered voucher scheme financed through a proportional wage tax. Chile’s system has a similar financing mechanism, but the distribution of benefits is less progressive, so that the net effect is in principle less redistributive. Argentina’s remodeled obras system is something of a half-way house: the financing base is similar, but the distribution of benefits in terms of the quality of insurance increases with income, but there is some implicit redistribution from richer to poorer obras. On the face of it, Brazil’s health insurance system serves less of a redistributive function that those of the other countries, to the extent that there is no earmarked tax dedicated to financing health insurance. This however highlights the limitations of examining the health sector independently of the general tax and transfer system. The taxes paid by higher-income individuals in Brazil are not reduced when they opt for private insurance: the real question is then how redistributive is the general tax base, an issue I have not addressed in this paper.

Finally, let me briefly mention the issue of non-insurable or systemic risks. In health, such risks are usually associated with disease epidemics and/or natural disasters. However, recognizing the links between economic opportunities and health status, it becomes apparent that adverse economic impacts deriving from macroeconomic fluctuations may have important effects on health needs and the performance of health insurance institutions. An important parameter in an analysis of these
links would be the rate at which health status declines in the face of negative income/consumption shocks, and the extent to which otherwise temporary macroeconomic shocks translate into permanent decreases in welfare for sub-populations. The ability of a health system to respond quickly to income-induced health shocks could be an important factor in reducing the effects of hysteresis on poverty.

**Bibliography**


Hart, Oliver, Andrei Shleifer, and Vishny (1997)


Larrañaga, Oswaldo (1997): “Reforms of the health sector in Chile,” mimeo.


World Bank (1994):

