INCENTIVES AND THE REFORM OF

HEALTH CARE SYSTEMS

PAOLO CARLO BELLI

LONDON SCHOOL OF ECONOMICS AND POLITICAL

SCIENCE

DEPARTMENT OF SOCIAL POLICY

THESIS SUBMITTED TO THE UNIVERSITY OF LONDON IN PARTIAL
FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF **DOCTOR
OF PHILOSOPHY**

UMI Number: U209544

UMI

Dissertation Publishing

UMI U209544

ProQuest

THESES

F

8563

PAOLO CARLO BELLI: INCENTIVES AND THE REFORM OF HEALTH CARE SYSTEMS

ABSTRACT

This thesis is a study of the reform of health systems from an international and an economic perspective. Its main unifying theme is to investigate the role played by incentives in the performance of health systems and their reform. In the first part, the thesis reconsiders the economic reasons that form the basis for public intervention in health markets, both in financing as well as in service provision. In fact, one of the key elements introduced with health reforms in the last few years has been greater competition in health insurance and provision, among private as well as public providers. It is thus interesting to start the analysis by revisiting the effects of competition in health markets on the basis of more recent contributions in microeconomic theory, our aim being to ascertain what would be the major deficiencies of unregulated markets, and to investigate into the impact of different public corrective measures. Chapter 2 looks at the effects of competition in the health insurance market and at the impact of different forms of public intervention to correct market failures. Chapter 3 presents a model of oligopolistic competition between two health providers, and it investigates the potential role of quality and/or price regulation as a means to extend coverage/improve quality beyond the point reached in correspondence to the market equilibrium.

Then, the thesis focuses on the new resource allocation, contracting mechanisms and payment systems for providers (RAP reforms) implemented over the last few years, within the public sector, or intended to discipline the relationship with health care providers. Chapters 4 gives an introduction to the RAP reforms, their justification and main components. Chapter 5 focuses on payment systems and on efficiency issues, while Chapter 6 on the equity consequences of RAP reforms.

Chapter 7 and 8 look at the health reforms implemented over the last decade in the former socialist countries. The evolution of health systems in those countries provides interesting lessons, illuminating the major weaknesses and limitations of the health reform model that has been prevailing and proposed world-wide over the last decade. Chapter 8 presents a qualitative study of the impact of the health reforms in Georgia, focusing specifically on the phenomenon of out-of-pocket payments, formal and informal, which currently are the prevalent source of funding for health in the region. A concluding chapter (Chapter 9) summarises some of the main findings of the thesis.

Ai miei genitori

Acknowledgments

During the years I have worked on this research, I have been posted at five different research institutions in the UK (the London School), in Italy (Bocconi University and University of Pavia), and in USA (World Bank, Washington, and Harvard School of Public Health, Boston). My warmest thanks go to all of you who have inspired, assisted, commented, challenged and critiqued my work.

In particular, I would like to thank:

Julian Le Grand, my supervisor, for giving me freedom and for trying to help me stay on track, for being concerned with the details and helping me to see the big picture

Jean Dreze, for introducing me to the issues of development, for being a source of inspiration, and for taking me to Palanpur.

Tom Bossert and Peter Berman, for introducing me to the issues of International Health, for sharing your knowledge, for your support, and for being explicit and controversial, but with a smile.

Fadia Saadah, Anabela Abreu and all the other World Bank staff I worked with, for their patience and gentleness, and for making me feel I might have understood one or two things.

Helen Shahriari, for valuable advice on collection and analysis of qualitative data, and for being a really good friend, always having time to listen, and assist.

George Gotsadze, for helping me to understand the Georgian health sector.

David Daniels and the DFID Health Systems' Resource Centre, for having brought me back to London, and for being a solid source of support.

Karan Madan and the others of the Boston gang, for being friends when I most needed them.

Silvia Melloni, for your patience and support, and for being my companion on the first part of the journey.

Carlo Belli e Giuseppina Teggi, my parents, for their loving support. This thesis is dedicated to them.

Table of contents

List of Tables

List of Figures:

List of Abbreviations

CIS = Commonwealth of Independent States[1]

CEE[2] = Central and Eastern Europe

CPI= Consumer Price Index

DRG = Diagnostic Related Group

ECA = East Europe and Central Asia

FSE = CEE+FSU= Former Socialist Economies

FSU = Former Soviet Union (same as CIS)

GEL = Georgian Lari -Georgian National currency

GDP = Gross Domestic Product

GP = General Practitioner

HA = Health Authority

IP = Informal Payments

LIC=Low-income countries[3]

MCH = Maternal and Child Health

MIC = Middle-income countries[4]

MoH = Ministry of Health

MoLHSA = Ministry of Labour, Health and Social Affairs of the Republic of Georgia

MQS=Minimum Quality Standard

OOP = out-of-pocket payments

PC = Price Cap

PCT= Primary Care Trust

PS = Payment Systems

RAP=Resource Allocation and Purchasing

THE= Total Government Health Expenditure

THE=Total health expenditure

TPE= Total private health expenditure

---

[1] It includes Moldova, Latvia, Lithuania, Estonia, Russian Federation, Azerbaijan, Georgia, Armenia, Ukraine, Belarus, Kazakhstan, Kyrgyz Republic, Uzbekistan, Tajikistan, and Turkmenistan.

[2] It includes Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Hungary, Macedonia, Poland, Romania, Slovak Republic, Slovenia, Yugoslavia, and F.R.

[3] Low Income Countries, as defined by the World Bank Atlas method, are countries with gross national income (GNI) per capita lower than $766.

[4] Middle Income Countries, as defined by the World Bank Atlas method, are countries with gross national income (GNI) per capita comprised between $766-$9385.

Chapter 1: Introduction: The Research Questions, Outline and Methodology

1.1 Questions and Outline

This thesis is a study of the reform of health systems from an international and an economic perspective. Its main unifying theme is to investigate the role played by incentives in the performance of health systems and their reform.[5]

In fact, incentive problems caused by imperfect information have been at the core of the economic research on health systems. There are two characteristic market failures in health. The first is the principal-agent problem associated with the fact that nurses and doctors have better information than patients as far as appropriate treatment is concerned, as well as at times differing interests in that treatment, both financially and in terms of appropriate effort. The second source of market failure concerns the health insurance market. Health insurance plays an essential role in health care. Outside of routine care, medical expenses are relatively large and occur randomly, precisely the sort of problems for which there should be a large demand for insurance. However, in most developing countries health insurance is simply not available to the vast majority of people. The reasons are the standard ones of adverse selection, cream skimming and moral hazard which, in general, are severe enough to prevent the efficient and equilibrate functioning of health insurance markets.

Because of the absence of an effective private insurance coverage, the potential market failures in the provision of medical services, and for other reasons, such as equity, governments have generally been prominent players in health financing and provision. The principal means of intervention have been either direct provision of services at highly subsidised rates, or the sponsorship of social insurance programs for paying private and public providers of services. However, the public acceptance of responsibility for health services does not eliminate all of the problems that prevented an efficient market from emerging in the first place. In industrialized countries, public provision,

---

[5] Health systems' organisation and performance are obviously influenced by social, cultural and historical factors, but we believe that "incentives matter", in the sense that different financial arrangements, regulatory structures and degrees of market exposure are likely to strongly influence health outputs and outcomes, in a

while effectively handling the lack of insurance coverage by charging low fees for expensive procedures, has been plagued by serious problems associated with incentives facing providers. As a result, quality of services has frequently been a major issue in public provision. While private providers, usually paid fee-for-services, have an incentive to over-treat, a problem made worse when insurance is available, salaried civil servants create the opposite problem of having insufficient incentives to provide conscientious care. Provider institutions, reimbursed according to their inputs and historical expenditure, are unaccountable to patients and isolated from other providers' competition, and thus have no incentive to improve quality of services or to use resources efficiently.[6]

These quality/efficiency problems because of market and government failures have been much more severe in low income countries (see Chapter 6), where government institutional capacity and consumers are weaker, and private providers are less developed. Furthermore, in these countries with few exceptions, public health systems have also largely failed on equity grounds, because publicly funded services have disproportionately benefited the better off.

Over the last two decades, in order to respond to some of the above flaws, several governments in industrialized as well as in poor and middle income countries embarked on broad health reforms. Such reforms became known as internal or quasi market, or managed competition health reforms, or, where the focus was on changing the criteria according to which financial resources would be distributed within the health system, resource allocation and purchasing (RAP) reforms (see Chapter 4). The American health economist Enthoven (1985, 1991) was the first to articulate the conceptual framework behind these reforms, followed by several other scholars at the beginning of the '90s (see, for example, Maynard, 1991, and Le Grand, Bartlett, 1993).

The United Kingdom was the first among West European countries to implement health reforms inspired by the quasi-market model. Almost in the same period, comparable reform proposals began to be discussed in other countries of continental Western Europe, such as the Netherlands and Italy, of Latin America, such as Chile and Colombia, in some of the former socialist countries of Eastern Europe and the Former Soviet Union, such as the Czech Republic, Poland, Georgia and Estonia, in a few countries in East Asia (Thailand), and in Africa (Zambia).

---

sense that abstracts from the particular cultural and social influences that characterize each country.

In fact, each of these countries implemented some different variation of a common reform model, adapting it from its previous institutional structure. For example, in designing the managed competition model, Enthoven considered a country (the USA) characterized by a plurality of health purchasers and insurers, competing with each other. This was the situation also in some countries of Latin America, such as Colombia. However, when the managed competition model was adapted to health systems characterized by universal health insurance coverage and a monopsonistic public purchaser of services, it took a different shape. However, a few core features are shared by all countries (see Chapter 4):

> Greater exposure to market forces (particularly in health provision), with the purchaser-provider split, and the creation of public-private competition.

> New criteria for allocating financial resources across regions, districts, purchasers[7] and providers[8] (RAP reforms) within the public sector.

> The extension of decentralized public agencies' (local governments, health purchasers and providers) degree of financial and managerial independence.

> (On occasion), reform and partial "liberalization" of the revenue collection component of the health system, with the creation of a plurality of social health insurance funds, greater use of co-payments from patients, greater role for private health insurance and for community-based schemes.

The research in this thesis is aimed at providing an analysis of the above reforms from an economist's perspective, focusing on the impact of the new incentive system and institutional framework on efficiency and on equity. It is divided in three main parts, closely connected to each other.

In the first part, the thesis revisits the economic rationale for public intervention in health markets, both in financing as well as in service provision. In order to do so, it first analyses the dynamics of competition in health markets and the characteristics of market equilibria, building upon recent contributions in microeconomic theory. Our aim is to describe equilibria in unregulated markets, both in provision as well as health insurance, and to investigate the impact of different public corrective measures. The central questions we pose in this first part are the following: "How would

---

[6] Similar problems have also been common in other parts of the public sector, such as education.

[7] The purchaser is an agency interposed between patients and providers, whose role is to "organize specific types of health care for a designated population (whether defined by geography, employment type or voluntary enrolment)" (Rice and Smith, 2000, p. 1). See Chapter 4.

[8] In the thesis, we refer to both individuals and institutions supplying health services as health providers.

1) the level of coverage in the private health insurance market, and 2) the quality/price of provision set by the private sector be characterized?" "What would be the impact of different government interventions?"

Chapter 2 looks at competition in the health insurance market, and analyzes the impact of different forms of public intervention to correct market failures. It also introduces in an accessible way some of the core features of the so called 'economics of information' models, the analytical framework developed in economics over the last twenty years to analyse market interaction in situations characterized by imperfect and asymmetric information among economic agents.

Chapter 3 focuses on the provider side. It presents a model of oligopolistic competition between two health providers, and it investigates the potential role of quality and/or price regulation as a means to extend coverage/improve quality beyond the market equilibrium level.

Understanding the dynamics of market competition in health insurance and in health provision would allow evaluating health policy interventions in a new light. The two chapters show that the impact of any health policy intervention, such as a demand or a supply side subsidy or an increase in the rate of co-payments or the imposition of a Minimum Quality Standard, depends on the characteristics of the equilibrium that prevails in the market before the public interventions, and on how that intervention impacts upon that equilibrium.

The second part of the thesis focuses on the internal market reforms implemented over the last two decades in several countries. The central question the thesis addresses in this second part is the following: "What are the core characteristics of the reforms that have been implemented, and what are their key consequences for equity and efficiency?" Chapter 4 introduces the reforms and to the conceptual framework we propose to assess their impact. Chapter 5 focuses on payment systems (PS) and on efficiency issues, and presents an original principal-agent model to investigate the properties of the second-best optimal contract between a purchaser and a provider of health services.

The first five chapters of the thesis are mainly theoretical. The market interaction in the health insurance and the health delivery markets, as well as the possible reaction of purchasers and providers to different resource allocation and payment systems have been analysed in the health economics literature either through principal-agent models or through monopolistic competitive models. The thesis will present both type of models, and build upon the more recent developments

in the literature to analyse specific aspects of the health market interaction in service delivery (in Chapter 3) and the payment system component of the reforms (in Chapter 5). The relative advantage of principal-agent theory is that it recognises and models explicitly the potential conflicts of interest between different actors or agents, emphasising asymmetry of information as the critical problem in the discipline of insurers and providers. This theory can also have both a positive content and a normative content, clarifying in which direction and to what extent different government interventions (including regulatory regimes, subsidies, etc.) are likely to modify the overall performance of the health system. The relative advantage of monopolistic competitive models is that, unlike principal-agent models, they explicitly consider the effects of competition among a plurality of health providers.[9]

The last three chapters of the thesis look at the health reforms implemented over the last decade from an empirical perspective, giving account of the existing evidence and presenting an original study on the impact of the reforms in the Republic of Georgia. These chapters illuminate the major weaknesses and limitations of the health reform model that has prevailed world-wide over the last two decades.

Chapter 6 reviews the evidence available from developing countries on the distribution of health and health benefits, clarifies what we mean by an equitable or more equitable allocation of health resources, and finally reviews our empirical and conceptual understanding of the equity impact of different reform allocation and purchasing (RAP) reform components. Chapter 7 presents an overview of the evolution of health indicators and health expenditures in Central and Eastern Europe and the Soviet Union over the last few years. Chapter 8 zooms on the reforms in a particular country, Georgia, and looks specifically at the constraint to accessibility of services posed by the raise of out-of-pocket (OOP) payments as the main source of revenue for health providers. Chapter 9, the final chapter, summarises some of the main results of the thesis.

While the agenda sketched above is certainly ambitious, it is hard to cut off the analysis without doing injustice to the complexity of the policy questions involved. If focused only on the market failures, the research may lead to excessive enthusiasm for public provision, as is sometimes demonstrated by reformers in the United States. If focused entirely on the problems of public provision, there will be a tendency to fall back on solutions of "letting the market handle it" as is still heard sometimes in Europe. If only the polar cases are examined, the research risks losing relevance.

---

[9] See, for example, D. Dranove and M. Satterthwaite (1992).

It is important to show the unifying policy problems --the difficulties in insuring against health risk and encouraging high quality, cost-effective care- before limiting the analysis to more digestible pieces. At the same time, it is important to show that the institutional arrangements experimented in the last two decades in several countries, based on the introduction of market incentives and mechanisms within the framework of a public health system, may lead to the emergence of new dilemmas and trade-offs. The thesis attempts to shed light on some such trade-offs and to be a guide for better health policy analysis and intervention.

1.2 Methodology

This is a thesis which aims to be policy-relevant and to be of interest not just for an academic audience, but for a broader audience of health policy experts. One of its original features is its ability to build up a coherent conceptual framework considering contributions from several theoretical (from health economics and microeconomic theory), policy-oriented, and empirical works.

In order to achieve this aim, my strategy has been to provide both survey papers and more rigorous theoretical contributions in the first part, and a more empirical analysis in the second part. To explain it through a metaphor, the survey papers should accompany the reader in an introductory "skating-tour" across the broad iced lake of the relevant literature concerning the issue at stake in each Chapter, while the models are holes in the ice, to go deeper in the analysis of specific issues. In fact, the "lake" of the literature relevant for this research proved to be immense. The hard-core economic literature alone comprises part of welfare economics, agency theory and its applications to the theory of regulation, incomplete contract theory and oligopoly analysis, and their application to agents' behaviour in health markets. More broadly, the lake encompasses all the health economics and policy-oriented literature that has investigated into the role of government intervention in health markets, from an equity and from an efficiency perspective, the literature that has dealt with equity, its measurement and its implications, plus the policy-oriented literature developed in each country and at the international level to analyse health reforms implemented during the last two decades. I wanted to give an account of at least some of this immense and diverse research edifice.

The two "holes in the ice" or theoretical models developed in the thesis build upon some of the more recent developments in the theory of regulation (Laffont and Tirole, 1993) in Chapter 5, and in industrial organization (vertical competition models) in Chapter 3.

On the empirical part, most of the information presented in the first six chapters is taken from other studies, but the evidence presented in Chapter 6 on the Republic of Georgia is original.

Chapter 2: Adverse selection and the Health Insurance Market

2.1 Introduction

This chapter investigates the way adverse selection influences health insurance market outcomes. Adverse selection represents just one of the sources of market failure that justify government intervention in health, as highlighted in the theory of welfare economics. Thus, before analysing the problem of adverse selection, the central theme of this chapter, it is useful to briefly outline also the other sources of market failure in health markets. The other key reason for government intervention, equity, is discussed extensively in Annex 2,1, while moral hazard is briefly introduced in Annex 2.2.

2.2 Sources of Market Failure in Health

The two fundamental theorems of welfare economics (Arrow-Debreu, 1952) are the starting point of the neoclassical economics analysis of health markets. The two theorems show that, if certain conditions are satisfied (convexity of the production set and price-taking behaviour by firms, convexity, continuity and monotonicity of preferences and price-taking behaviour by consumers, complete markets and perfect information, absence of externalities and public goods[10]) competitive equilibria are Pareto efficient [11] (1st theorem), and that any Pareto efficient equilibrium may be reached as a competitive equilibrium by appropriately redistributing initial endowments across individuals through ex-ante lump-sum transfers (2nd theorem)[12].

Health care markets violate some of the conditions necessary to achieve Pareto optimal equilibria for the following reasons:

> ➢ First, there are specific health services, such as sanitation and health promotion, which are non-excludable (free riders cannot be excluded from consumption) and non-rival (use by

---

[10] See Varian, 1992, p. 326.
[11] A given allocation of resources is Pareto optimal (strong Pareto optimality) if it is not possible to further improve the welfare of any individual without decreasing that of another one.
[12] The terminology used in general equilibrium theory refers to Walrasian, rather than competitive equilibrium. All is requested to characterize the market interaction is that firms and consumers are price-takers.

one consumer does not preclude use by others). Hence, they would not be offered by the private for-profit sector: in the language of welfare economics, they are public goods. Second, there are services, such as immunisations, which are in principle excludable and rival, but entail large positive externalities. By receiving immunization, an individual reduces the probability of contagious infection not only for herself, but also for others. This means that the social benefits of immunization and other preventive interventions against communicable disease are larger than the private benefits, and that, if not subsidized, private for-profit providers would supply a sub-optimal quantity of services [13].

➢ Transactions in health markets are largely decided by supply. Physicians dispose of better information than patients do concerning the appropriate diagnostic and treatment procedures, as well as having differing interests from patients, both financially and in terms of appropriate effort.

➢ Imperfect and asymmetric information characterize the health insurance market. Adverse selection, moral hazard, and the impossibility of writing complete, long-term contracts can prevent the efficient functioning of the market.

Note that according to neoclassical economics the above sources of market failure are necessary, but not sufficient conditions to justify public intervention[14]. Moreover, once an agreement is reached over the relative advantage of public intervention, one must still solve the problem of finding the mix of public policies that in each circumstance are most appropriate to address the particular source of failure that prevents markets from achieving efficient outcomes (see also Annex 6.4, on the limits of redistribution). Two broad dimensions must be specified:

a) The first concerns the choice of an appropriate mix of governmental instruments. These include: 1) provision of information; 2) taxation/subsidization of the private sector; 3) regulation; 4) direct investment. As we shall see in this and the following chapter, each of these different forms of interventions produces a distinct impact on market equilibria.

b) The second dimension, for those services that the government chooses to directly provide, concerns the design of an appropriate organization system through planning of human and

---

[13] Under-provision problems are more severe when people lack information on the benefits of such interventions (see analysis in the text, below).

[14] It has been proved that apparently sub-optimal private market equilibria may in fact be constrained Pareto-efficient, such that there is no room for improvement, once the same information constraints that limit markets are imposed also upon public action (Harris and Townsend, 1981). This is the case, for instance, within a pure moral hazard context in the health insurance market. In macroeconomics, similar conclusions that restrict the possibility of improving upon macroeconomic equilibria are reached, once we abandon the hypothesis that individuals' expectations are myopic or slowly adapt to policy changes (see Sargent and Wallace, 1985).

physical infrastructure, and a mix of information and financial instruments.

In other words, proving the possibility of a Pareto-improving public intervention in the market is the first step of a complicated policy analysis exercise. It is then necessary to define how any corrective public intervention could and should be articulated.

This Chapter will illustrate in detail these different options and their impact in the case of insurance markets with adverse selection. We shall show that in general some positive level of cross-subsidies across different risk groups characterizes second best equilibria[15], and that the free market equilibria may fail to sustain such cross-subsidization. Thus, adverse selection can be used as the justification for the provision of compulsory universal public health insurance. The claim is that, due to adverse selection, the government would in any case bear a significant share of total health costs even if it chose to take care only of those individuals and risks who are not insured privately. By forcing everybody in the same insurance pool, the government could guarantee stability of health financing and make the cross- subsidies across risk-groups more transparent. In fact, the government has many different options to correct market equilibria, and maintain a positive level of cross-subsidy among different risk groups. The options available are to provide full or partial public insurance with or without the possibility of complementing it with private insurance, and with or without the possibility of opting out of the public scheme. Alternatively, the private insurance market could be regulated, for instance through the imposition of a standard contract or of a minimal insurance guarantee. As we shall see in the second part of this chapter, the above interventions entail different results in terms of efficiency.

2.3 Adverse Selection

Adverse selection is defined as the strategic behaviour by the more informed partner in a contract against the interest of the less informed partner(s). In the health insurance market it is relevant because each individual chooses among the set of contracts offered by insurance companies (in the following referred to as "plans") according to his/her expected probability of using health services. In brief, those who foresee an intense use of health services will tend to choose more generous plans

---

[15] In other words it is in the low risk group's interest to provide some form of subsidy in favour of the high-risk individuals as a means to achieve a degree of coverage closer to their preferred full coverage contract (see analysis in the text, section 2.7).

than those who expect a limited use of them. In the extreme case, for each premium and extent of coverage set in any plan, those who will decide to purchase it are only those who expect to sustain an amount of health expenditure greater or equal to the premium they need to pay. Then, without adjustments to prevent adverse selection, insurance companies would end up with a loss on each customer, regardless of the premiums they charge.

However, insurance companies do anticipate adverse selection and devise contract offers in order to screen individuals, so that they can charge each potential customer a premium commensurate to his/her expected cost of coverage. [16] These screening strategies by insurers generally hinder the achievement of an efficient risk pooling across individuals. The market can be trapped in a sub-optimal equilibrium, where insurance companies offer only incomplete coverage, in spite of the fact that there is a positive demand for more comprehensive insurance.

There is a growing body of evidence that suggests that adverse selection is an important phenomenon in health insurance markets. Cutler writes (1996, p.30): "Almost all health insurance systems where individuals are allowed choice of insurance have experienced adverse selection. Medicare enrolees who choose managed care[17] are healthier than ...[those] who do not. The Federal Employees Health Benefits Program...has adverse selection between more and less generous policies. The spread in premiums between more and less generous policies is 68% greater than benefits alone would dictate...And almost every large firm that has encouraged employee choice has found the cost of the most generous policies increases sufficiently rapidly that these policies are no longer viable" (this last phenomenon is known in the literature as a "price death spiral" and it is a consequence of adverse selection; see below). It is expected that the phenomenon of adverse selection will become more severe in the USA and other countries, if the current trend continues, which is characterized by increasing competition in the health insurance market and by the diffusion of new employment-related schemes where individuals are brought to face the true marginal cost of their health coverage. [18]

---

[16] This "screening" strategy is even more critical to success in the market whenever there is regulation in place that does not allow health premiums to reflect individual risk (premium rate restrictions), or does not permit to acquire information on potential customers' health condition before finalizing contract offers.

[17] Managed care plans impose stricter controls and restrictions over use of health services than traditional indemnity plans.

[18] In the past, employers would pay a large share of the premiums. Increasingly, employees are offered lump-sum transfers for health insurance and they face almost entirely the relative marginal costs of alternative plans.

In this chapter we will first illustrate the problem of adverse selection (AS) by way of two examples, and then present a selective survey of the theoretical literature that has described the impact of the phenomenon of AS on market equilibrium. In the second part, the paper analyses different policy options to correct market equilibria. We shall try to provide for the first time a unitary representation of a set of concepts that have developed piece-meal over a period of more than twenty years. Given the aim of our survey and the broadness of the existing literature, our attention is focused on giving an intuitive understanding of the main results, rather than in presenting them in an analytical rigorous form. Whenever possible, we will make use of diagrammatical proofs. Chapter 5, which adopts a similar analytical framework of a principal-agent model with asymmetric information to study the purchaser-provider relationship, will describe in detail the mathematical properties of this family of models.

2.4 Two Examples

The first example builds on a similar case presented by Cutler and Zeckhauser (1997). Consider two health plans, a generous (G) and a moderate (M) plan, offered in a particular market composed of two types of individuals, high-risk and low-risk, each group making up 50% of the entire population. Suppose that the cost of treating individuals under the two plans, and their gains in benefit from G vis-à-vis M, are as follows:

Table 2.1: Resource Cost and Benefit from Health Insurance Coverage (D=Denarius)

	Resource cost of coverage	Resource cost of coverage	Benefit difference: generous – moderate (money equivalent)
	Moderate Plan	**Generous Plan**	
Low-risk	40 D	60 D	15 D
High-risk	70 D	100 D	40 D

Let us assume that insurance market is perfectly competitive so that in equilibrium premiums must be equal to expected costs. First note that the first best or full-information (FB) equilibrium in this example would see the high-risk individuals enrolled in the generous plan for a price of 100 D (G-

M=difference in benefits=40>30=difference in costs), and the low-risk individuals in the moderate plan for a price of 40 D (G-M=difference in benefits=15<20=difference in costs).

However, the above FB equilibrium is not sustainable with incomplete information, when insurers only know that each potential customer can be either low or high-risk with equal probability[19]. At a price of 40 both groups would buy the moderate plan, which would start making losses (if both groups purchase M, is expected cost of coverage is 55D (40*0.5+70*0.5).

Now consider the following situation. Initially a unique plan, the generous one, is offered in the market. If the market is competitive in equilibrium such plan must break even, and it would thus be offered to everybody for a price of 80D (60*0.5+100*0.5). Then, let us assume that the moderate plan is devised and offered for a price of 64D, which is low enough to attract low-risk individuals. All low-risks would switch to the new plan (they can save 16 in exchange for a benefit loss of 15), while the high-risk individuals would stay with the generous plan (their net benefit loss from switching would be equal to 24=40-16).

However, once the low risk have moved away from G, at the original price equal to 80D the plan, now burdened with all high-risk individuals, would become unprofitable, while M at a price of 64D would be making a positive profit. Thus, the forces of competition would lead to an increase in the price of G and a decrease in the price of M. When the price differential between G and M exceeds 40D (given the above assumptions, it will eventually do so, because in equilibrium premiums must reflect relative expected costs), all high-risk individuals will also switch to M. Then G, the generous plan, is terminated, and, when this happens, if the price of M is below 55, M would also become unprofitable, because once it shoulders all risk types its price must exceed 55 (=70*0.5+40*0.5). When the 'pooling equilibrium' (see below) in M is restored, new opportunities to undercut M, by offering yet 'more' moderate plans that attract low-risks, may arise. The market may be characterized by chronic instability.

Now, consider the equilibrium that the market would reach by changing Table 2.1's figures for net benefits as follows:

---

[19] Insurance companies can also compute the expected cost of coverage for each of the two risk-groups separately and the expected costs of the plan that pools together both groups.

Table 2.2: Resource Cost and Benefit from Health Insurance Coverage (D=Denarius)

	Resource cost of coverage	Resource cost of coverage	Benefit difference generous – moderate (money equivalent)
	Moderate Plan	**Generous Plan**	
Low-risk	40 D	60 D	25 D
High-risk	70 D	100 D	60 D

Both groups are made more "risk averse" in this second example: they both value being in the generous plan more than in the previous example. As before, suppose that insurers do not know each individual's risk type. It is clear that the full-information equilibrium, which would in this case see both risk groups purchasing the generous plan for a price of 60 (low-risk) and 100 (high-risk), cannot be an equilibrium with incomplete information. Moreover, starting from the same initial situation considered above (generous plan offered for a price of 80D), it is profitable for low-risks to switch to the moderate plan as long as it is offered for a price lower than 55D (which is the price where the amount low-risk individuals save by switching to M, 80-55 = 25, is equal to their net benefit loss from switching). Again, when low-risk individual abandon the generous plan, its price will need to rise up to 100D, which is the average cost to cover high-risks. At the same time competition will lead to a further decrease in the price of the moderate plan, down to 40D (=expected cost of coverage of low-risks). Unlike the previous example, however, a price differential of 60D (100-40) is not sufficient to induce high-risks to switch to the moderate plan (in fact, they are just indifferent between switching to M and maintaining G). Thus, in this case the situation in which high-risks pay 100 for full insurance and low-risks pay 40 for incomplete coverage is a stable market equilibrium (a 'separating equilibrium'; see below). With incomplete information, low-risks are worse off than in the full information equilibrium, as they obtain only partial insurance. The market "sorts" out low-risks from high-risk individuals by offering plans with less than optimal coverage. Low-risks would be willing to pay an additional 25D to get full coverage (and the cost of the additional coverage for them would only be 20D in case of perfect information); obviously, this option is not available to them in case of incomplete information, because by purchasing G they would be 'confused' with high-risks, and would have to pay more than their marginal cost of coverage.

2.5 Rothschild-Stiglitz and Equilibrium in Competitive Insurance Markets

Rothschild and Stiglitz (1976) were the first analytically to investigate the problem of adverse selection in the insurance market. We begin our survey of the theoretical literature by presenting a detailed summary of their model, which is then going to be used to illustrate all the subsequent developments in the literature.

Rothschild and Stiglitz (1976) analyse a situation where individuals are subject to exogenous shocks to their wealth. We can imagine that such shocks (named "accidents") can be provoked by diseases that prevent individuals from continuing their normal activities.

Individuals' wealth prospects depend on whether or not any insurance is available in the market: without insurance, wealth is equal to $W_1=W$, if no 'accident' occurs, and to $W_2=W-d$ if it occurs, where "d" is the amount of the expected wealth loss. If insurance is available, in the simplest case one can assume that insurance companies offer an indemnity contract or plan, with a reimbursement equal to $\alpha^{\wedge}_2$ if an accident occurs in exchange for a premium (paid in any case) equal to α_1. Individuals' wealth with insurance becomes respectively: $W_1 = W - \alpha_1$ if no accident occurs, and W_2 $= W - \alpha_1 + \alpha^{\wedge}_2 - d = W + \alpha_2 - d$, where $\alpha_2 = \alpha^{\wedge}_2 - \alpha_1$, is the net indemnity, if an accident occurs. Suppose an accident occurs with probability p. Then, if insurance is available one can represent individuals' preferences for wealth in the two states in the following way[20]:

$$V(p, \alpha_1,\alpha_2) = (1-p)U(W-\alpha_1) + pU(W + \alpha_2 - d) \qquad (2.1)$$

Given p, the probability that an accident occurs, each individual chooses the level of insurance that maximizes $V(\cdot)$ with respect to $\{\alpha_1,\alpha_2\}$. Rothschild and Stiglitz (hereafter, R.-S.) assume that individuals are risk averse, i.e. $U''(\ ) < 0$ [21], and that they cannot influence the amount of the indemnity, nor the probability of obtaining it (no moral hazard).

---

[20] The expression in the text writes utility as a linear function of probabilities. In order to do so, the description of preferences under uncertainty must satisfy a set of conditions. See: Varian (1992, pp. 172-176).
[21] Thus, $V(\cdot)$ being a linear combination of concave functions, is quasi-concave.

On the supply side of the market insurance companies are considered risk-neutral and only interested in expected profits. A contract offer C_i consists of a bundle $\{\alpha_1,\alpha_2\}$ containing a specific indemnity ('amount of insurance') that the individual can buy in exchange for a specific premium (in the diagrams that follow we will refer interchangeably to contracts C_i or to $\{\alpha_1,\alpha_2\}$, the net premium-indemnity pair which identifies each contract). R-S assume that the insurance market is perfectly competitive, which implies that in equilibrium expected revenues from premiums are equal to expected costs from indemnity payment, but at the same time that individuals can buy at most one insurance contract, thus recognizing that insurance companies are able to ration the extent of insurance coverage [22].

The expected profit associated with a contract to an individual with probability p of incurring an accident which would trigger the payment of an indemnity equal to $\alpha^\wedge_2$ is the following:

$$\pi\,(p,\,\alpha_1,\alpha_2) = (1\text{-}p)\,\alpha_1 - p\,(\alpha^\wedge_2 - \alpha_1) = (1\text{-}p)\,\alpha_1 - p\,\alpha_2 \qquad (2.2)$$

The equilibrium concept R-S adopt is that of Nash-Cournot, where each agent maximizes his/her objective function independently of other agents' reaction, and the equilibrium set of contracts is defined by the following conditions:

I. Customers maximize expected utility.

II. No contract within the equilibrium set entails negative expected profits.

III. No contract outside the equilibrium set, if offered, would make a positive profit.

Finally, R-S assume that, when deciding to sign a contract, individuals know "p", the probability that the accident occurs, while insurance companies do not.

2.5.1 Equilibrium with Identical Customers

Let us first consider the equilibrium with identical customers. Figure 2.1 represents on the horizontal axis wealth if no accident occurs, and on the vertical axis wealth if it occurs. Situations of full

---

[22] Later literature pointed out at the inconsistency of the two above assumptions. In theory each consumer should be able to buy unlimited quantity of a given good offered under competitive conditions at the prevailing market prices. R-S's assumption of quantity constraints over the available extent of insurance coverage would be more compatible with the assumption that the market is oligopolistic.

insurance correspond to points on the bisetrix, while situations of incomplete insurance lie to the right of the bisetrix (where $W_1 > W_2$).

Consider a representative individual, whose expected wealth in the two states without insurance is represented by point E. The individual is exposed to an accident with probability p, and in that case she/he sustains a loss equal to d. Each point to the "North-West" of E can be reached by purchasing a specific insurance contract, uniquely identified by a premium α_1 and a net indemnity α_2 in case the accident occurs. Segment EF represents the zero profit, or 'actuarial (fair) odds line. If any insurance company accepted to trade wealth in the two states with the individual at a rate equal to the slope of EF ($d\alpha_2/d\alpha_1 = (1-p)/p = -dW_2/dW_1$)[23], it would gain a zero expected profit, because expected revenue ($(1-p)\,\alpha_1$) would be equal to expected cost ($p\alpha_2$). By contrast, starting from point E, all points to the "South-West" of EF would entail a positive profit, and all points to the "North-East" of EF would result in a loss for the insurance company.

Thus, the above conditions I, II and III constrain equilibrium contracts to belong to the set along the 'zero profit' line, where $\pi(p, \alpha_1, \alpha_2) = (1-p)\,\alpha_1 - p\alpha_2 = 0$[24].(**Assumption 1**)

Figure 2.1 also represents individuals' preferences concerning combinations of wealth in the two states through a map of indifference curves. Given risk-aversion, the indifference curves are convex [25]. Given any indifference curve, all the points to the "North-East" of it entail higher utility, and all the points to the "South-West" entail a lower utility.

Figure 2.1: Equilibrium with Identical Customers
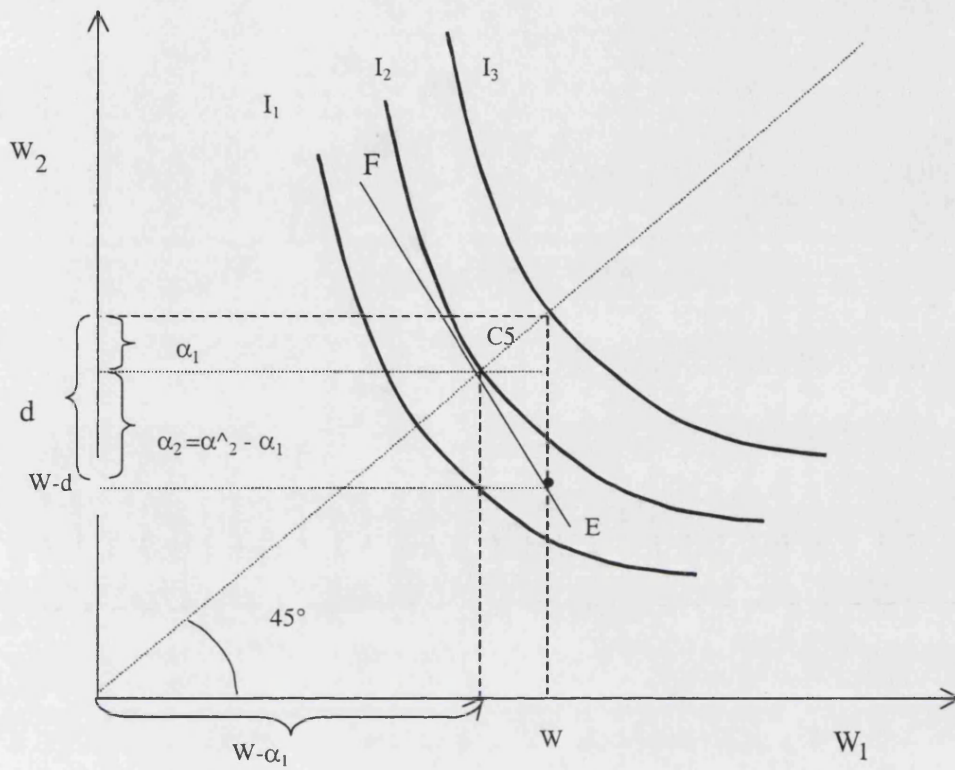
The equilibrium contract lies in correspondence to the highest indifference curve compatible with the budget constraint EF (point C_5 in Figure 2.1). In correspondence to C_5 the slope of the indifference curve is equal to the slope of EF. From the equation of the indifference curve:

$$(1-p)[dU(W_1)/dW_1]dW_1 + p[dU(W_2)/dW_2]dW_2 = 0, \tag{2.3}$$

If we denote: $dU(\ )/dW_1 = U'(W_1)$ and $dU(\ )/dW_2 = U'(W_2)$, the slope of the indifference curve is equal to:

$$dW_2/dW_1 = [U'(W_1)/U'(W_2)][(1-p)/p] \tag{2.4}$$

The tangency condition entails:

$$[\frac{U'(W_1)}{U'(W_2)}][\frac{(1-p)}{p}] = \frac{(1-p)}{p} \Rightarrow U'(W_1) = U'(W_2) \Rightarrow W_1 = W_2, given\ U''(\cdot) < 0 \tag{2.5}$$

Thus, given that individuals are risk averse and insurance companies are risk neutral, the first best is characterized by full insurance[26].

2.5.2 Equilibrium with Two Classes of Customers

Let us now consider a market consisting of two groups of customers[27]. They are characterized by the same utility function for wealth in the two states, $U(.)$, but by different accident probabilities.

- high-risk individuals, with probability of accident $= p^h$

- low-risk individuals, with probability of accident $= p^l$, with $p^h > p^l$

Let the percentage of high-risk individuals be equal to λ^{RS}. The average probability of accident is then equal to: $p^{EXP} = \lambda^{RS} p^h + (1 - \lambda^{RS}) p^l$

it is possible to distinguish between two types of equilibria:

- pooling equilibria, in which both risk groups purchase the same insurance contract. In any pooling equilibrium, the zero profit condition (Assumption 1 above) must hold across all the individual types:

$$(1-p^{EXP})\alpha_1 - p^{EXP}\alpha_2 = 0 \qquad (2.6)$$

- separating equilibria, in which different risk-groups choose different contracts. In any separating equilibrium, Assumption 1 above requires that each of the separate contracts yield zero expected profits. In the case of just two risk groups:

$$(1-p^h)\,\alpha_1 - p^h\alpha_2 = 0, \text{ and } (1-p^l)\,\alpha_1 - p^l\alpha_2 = 0 \qquad (2.7)$$

In Figure 2.2 we denote with the letter L and H the indifference curves and zero profit lines relative to respectively low-risk and high-risk individuals. Note that the slope of the zero profit line for low-risk individuals (L, with a slope equal to $(1-p^l)/p^l$) is steeper than that relative to high-risk individuals

---

[26] In fact, if the insurance premium were set at a higher level than the actuarially fair premium, the extent of the insurance coverage chosen by individuals would be lower (incomplete coverage). However, under R-S's hypothesis of free entry in the market, any plan with a premium above the "fair-odds" level would be undercut by competition until the zero-profit equilibrium is reached.

[27] We will frequently refer to them as risk-types or risk-groups.

(H, with a slope equal to $(1-p^h)/p^h$), because the probability of the accident occurring and the insurance company paying the indemnity is lower for low-risks. From equation (2.4) note also that the slope of the low-risks' indifference curve passing through each point $\{W_1; W_2\}$ is steeper than that of the high-risks' indifference curve through the same point. In other words, each indifference curve of high and low-risk individuals can intersect only once ("single-crossing property"). Finally, we have replicated in Figure 2.2 segment EF from Figure 2.1, which should now to be interpreted as the zero profit line for pooling contracts. Contracts on segment EF are the best individuals can hope to receive from insurers when they are all purchasing the same contract.

Figure 2.2: Equilibrium with Two Classes of Customers
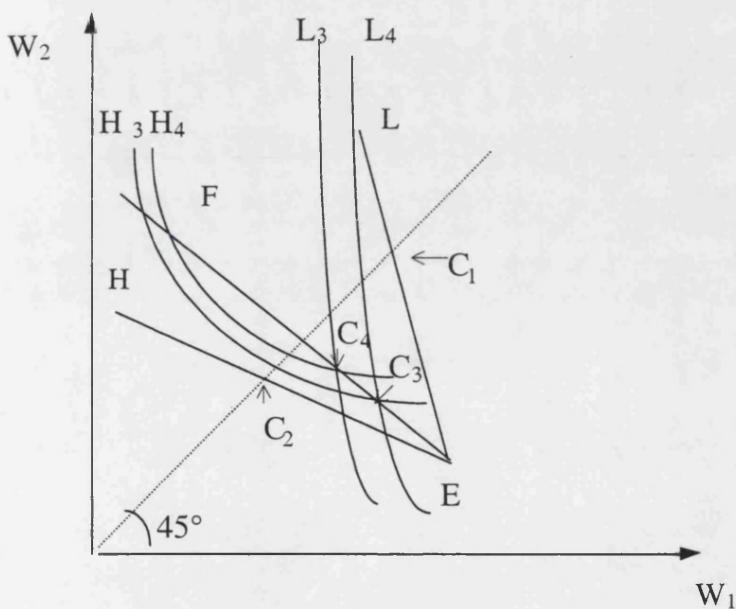
The first result R-S prove is that, when there are two different risk groups and information is incomplete, first best equilibria are no longer sustainable, as they violate the high-risk group incentive-compatibility (IC) constraint (for a detailed explanation of the meaning of the IC constraint, see Chapter 5). The IC constraints entail that, whenever different contracts are offered to each risk-type, they must be devised so that each individual would prefer the contract specific to its own risk-type vis-à-vis the contract set for any other risk-type. No one can be cheated or forced to buy a contract different from the contract she/he prefers among those available in the market. In our insurance market example there are just two risk-groups or types, high and low risk, so, if we denote

by C^h and C^l the set of contracts meant respectively for the high-risk and the low-risk groups, the I.C. constraints can be written as follows:

$$V(p^h, C^h) \geq V(p^h, C^l) \text{ and } V(p^l, C^l) \geq V(p^l, C^h) \tag{2.8}$$

As Figure 2.2 above shows, first-best equilibria violate the first of these constraints. If both contracts C_1 and C_2, the first-best contracts, are available in the market, all individuals would clearly prefer C_1 to C_2, because C_1 still provides full insurance coverage and greater wealth in both states. C_1 would then be unprofitable, because only contracts on segment EF, such as C_3 and C_4, or contracts to the 'south-west' of EF can be sustained as pooling equilibria.

The second, important result that R-S graphically prove is that there cannot be a pooling equilibrium.

Figure 2.3: Impossibility of a Pooling Equilibrium
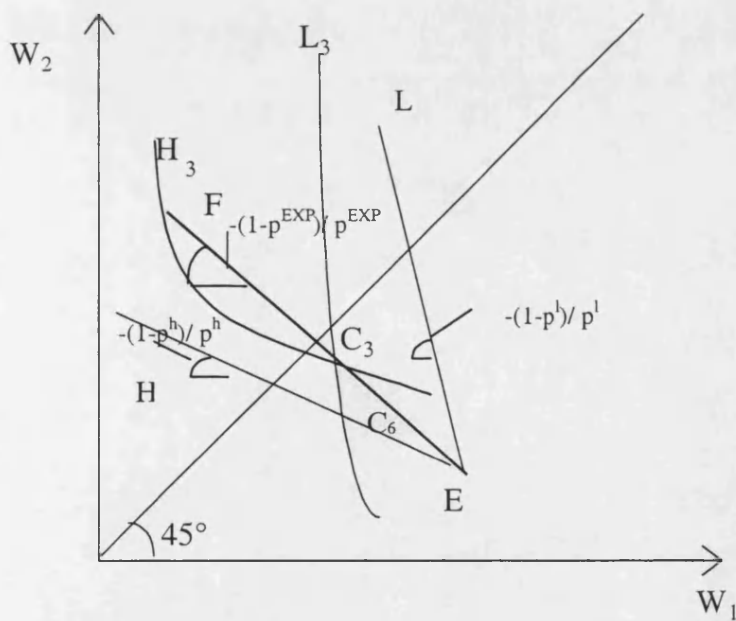
As Figure 2.3 indicates, any contract on segment EF, such as contract C_3, can always be upset by a new contract offer in the shaded area, such as contract C_6. All low-risk individuals are induced to purchase the new contract, which lies on a higher indifference curve for them, while high-risk

individuals would continue to purchase C_3. The latter contract, however, when burdened with only high-risk individuals, would become unprofitable and would be terminated[28].

The only possible equilibrium with different risk-types is a separating equilibrium, identified in Figure 2.4 by contracts C_2 and C_7. As Figure 2.4 indicates, when these two contracts are offered, high-risk individuals purchase contract C_2, characterized by full insurance, while low-risk individuals purchase contract C_7. Contract C_7 is the preferred contract by the low-risks among all contracts that respect the high-risk group's IC constraint $(V(p^h, C_2) \geq V(p^h, C_7)$[29]. Both contracts break even (contract C_2 lies on the high-risk zero profit line, and C_7 lies on the low-risk zero profit line). Thus, contracts C_2-C_7 satisfy conditions I, II, and III above, and they are a separating equilibrium.

Note that in any separating equilibrium the low-risk group receives incomplete insurance (C_7 lays to the right of the bisetrix), similarly to the situation described in the numerical example in Table 2.2. Note also that point C_7 is not at the tangency point between any indifference curve for low-risks and their budget constraint. When R-S published their seminal work economic theory still lacked a solid foundation for welfare analysis under conditions of imperfect information, but they had the intuition that the separating equilibria reached by the insurance market entailed a welfare loss. Low-risk individuals obtain less than optimal coverage while high-risk individuals do not improve their situation with respect to the first best. It is as if high-risk imposed a negative externality on low-risk. R-S also proved that the separating equilibrium is in general[30] Pareto inferior to a set of contracts characterized by some positive cross-subsidy from low-risk to high-risk (Rothschild and Stiglitz, 1976, pp.644-645), thus anticipating a key result which will later on be rigorously proven on the basis of second best analysis (see section 2.7).

---

[28] The impossibility of pooling contracts derives from the "single-crossing" property.

[29] In fact, in correspondence to the separating equilibrium C_2 C_7 the high-risk incentive compatibility constraint is binding, which means that high-risks are indifferent between C_2 and C_7. Given C_2, any contract along the low-risk group zero profit line above C_7 would be purchased by both groups and would therefore yield negative profits (when both groups purchase the same contract, such contract must lie on segment EF).

[30] The result does not hold if the percentage of high-risk people is high enough.

Figure 2.4: Separating Equilibrium with Two Risk Types
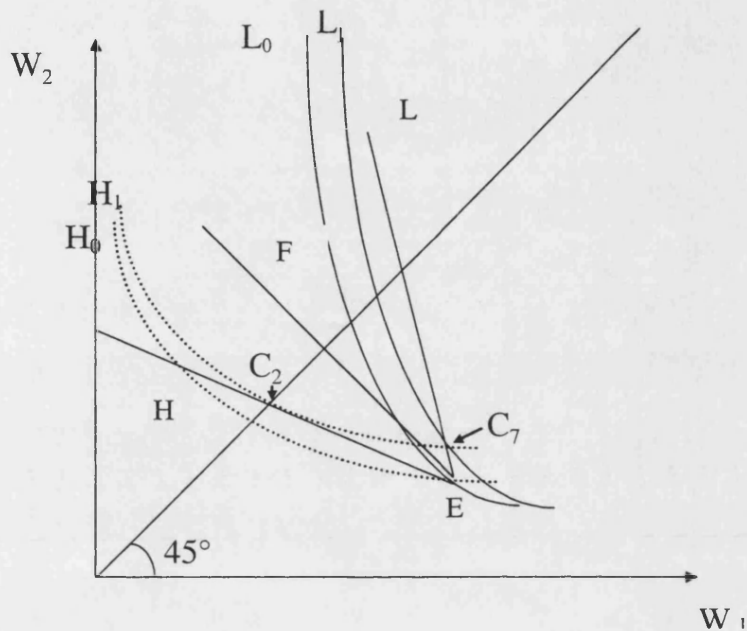
R-S also proved that a separating equilibrium may fail to exist, in which case the competitive insurance market has no equilibrium (as previously discussed -see Figure 2.3- any contract that pools both risk groups together, cannot be a stable equilibrium). This occurs when the cost of pooling for low-risk is not significant, while the cost of separating and thus accepting incomplete insurance is high, which is a situation that can arise under the following circumstances:

➢ When low risk individuals are characterized by a high degree of risk aversion.

➢ When the two groups are characterized by a similar probability of incurring the event for which they demand insurance.

➢ When there are few high-risk individuals.

Figure 2.5 illustrates a situation in which a separating equilibrium fails to exist:

Figure 2.5: Impossibility of a Separating Equilibrium
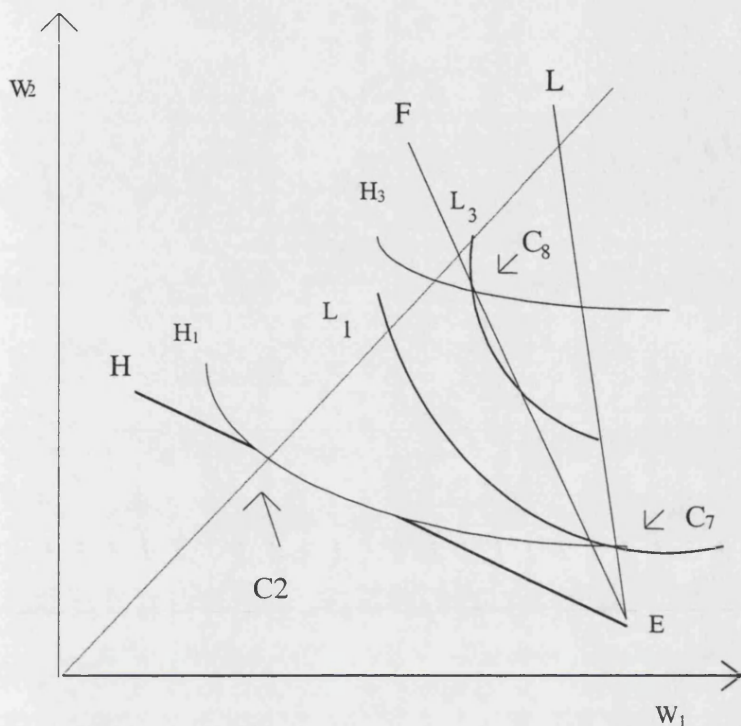
In the figure above, the potential separating equilibrium where all high-risk purchase C_2 and low-risk individuals purchase C_7 is upset by a contract that attracts both risk groups. Consider, for example, contract C_8, which lies on segment EF and so it is sustainable when both groups purchase it. Figure 2.5 indicates that C_8 is preferred by both groups to contracts C_2-C_7.

Graphically, one can see that the potential separating equilibrium is unstable whenever at least a sub-segment of EF, the pooling contracts zero profit line, lies above L_1, which represents the low-risk indifference curve in correspondence to the separating equilibrium[31]. Analytically, one can prove that there exists a threshold minimal percentage of high-risk individuals equal to λ^{RS}, which is a function of low-risks' degree of risk-aversion and of the risk differential across risk-types (p^h-p^l), that guarantees the stability of the separating equilibrium in the R-S model.

---

[31] Thus, the condition that must be satisfied for the separating equilibrium to be stable is that the low-risks' indifference (L_1 in the above figure) curve must be 'steeper' than the zero profit pooling line EF in correspondence to C_7.

2.5.3 Final Remarks on the R-S Model and Introduction to Subsequent Developments in the Literature

R-S describe the market interaction between potential customers and insurers in a situation of information asymmetry, where insurance companies do not know individuals' likelihood to fall in the situation (the 'accident in R-S's terminology) against which they demand to be insured. In order to improve available information on customers' risk, insurance companies can diversify plans and offer less expensive contracts characterized by limited coverage. Such contracts tend to be preferred by lower risk groups, and they can lead to separating equilibria. However, these equilibria may be sub-optimal, or/and unstable. The R-S's result is robust to changes in hypotheses as long as individuals with different risk properties differ in some characteristics that can be linked to their purchase of insurance and, somehow, insurance companies can discover that link.

R-S's analysis can be well applied to health insurance markets, where patients are characterized by private information over their health status and this information influences their decision to buy insurance. Younger and healthier individuals, and all those who feel it is unlikely they will become ill, would not value comprehensive insurance coverage as much as those who believe they would make more intense use of health care services. Insurance companies, in turn, know that patients' purchasing decision is influenced by their (perceived) likelihood of utilizing health care services, and offer 'cheaper' contracts characterised by incomplete coverage –for example by expanding the list of excluded conditions- in order to screen healthier individuals ('the good risks'). Once the 'good' risks are out of more comprehensive insurance schemes, the latter remain burdened exclusively with individuals affected by chronic disease or likely to undergo costly treatment, and their relative price tends to rise. In extreme circumstances, over time these more 'generous' health insurance schemes may remain accessible only to a thin, wealthier minority of the population willing to pay ever-increasing premiums, or be altogether terminated.

2.6 Equilibrium Refinements

Building on R-S's contribution, one stream of microeconomic literature, labelled as "equilibrium refinements", changed the hypotheses underlying firms' or individuals' behaviour in order to overcome R-S's puzzling result that in certain circumstances no equilibrium may exist in the insurance market.

R-S assumed that both agents, insurers and insured, follow a myopic (Nash) behaviour, which means that neither of them anticipates the other's possible reaction when deciding their strategy. Wilson (1976) was the first to remove this hypothesis and suppose instead that insurance companies behave strategically, in accordance with the analysis of firms' behaviour in oligopolistic markets. Wilson gives the following characterization of the equilibrium set of contracts (compare them with conditions I,II, and II in section 2.5):

I. Customers maximize expected utility.
II. No contract in the equilibrium set entails negative expected profits.
III. There is no potential new contract offer outside the equilibrium set that is profitable, when all loss-making contracts are withdrawn from the market as a result of the new contract offer.

Thus, unlike R-S, Wilson imposes that any deviation from the equilibrium must continue to be profitable even after all non-profitable contracts are discontinued.

A second contribution to the literature of "equilibrium refinements" is that by Grossman (1979), who makes the hypothesis of "dissembling" behaviour by high-risk individuals. According to Grossman, all potential health insurance customers know that in equilibrium any loss-making contract will be withdrawn from the market. Therefore, when high-risk individuals submit their application for insurance, they self-restrain their choice between the set of contracts chosen by low-risk and the set of contracts which would not be unprofitable even if chosen only by high risk (which in our proceeding analysis was identified with contract C_2 characterised by full insurance). In other words, high-risk 'dissemble' as low-risk by choosing the same contract low-risk choose, as long as this allows them to enjoy a level of utility higher than in C_2. By anticipating insurers' screening strategy, they realize that any other strategy would ultimately lead them to be all together excluded from the market.

Under Grossman and Wilson's hypotheses the same set of equilibria are selected. The R-S separating equilibrium holds when the proportion of high-risk λ is greater than a certain threshold λ^{RS}[32]. However, when $\lambda < \lambda^{RS}$ the pooling equilibrium preferred by low-risks becomes a stable equilibrium. When the percentage of high-risk individuals is small enough, low-risk prefer to cross-subsidize contracts for high-risk in the pooling equilibrium rather than to accept a lower level of

---

[32] Where λ^{RS} is the same percentage of high-risk individuals that guarantees the existence of the separating

coverage in the separating equilibrium. Thus, the pooling equilibrium preferred by low-risk becomes stable (at the point of tangency between segment EF and a low-risk group's indifference curve L_4 in correspondence to contract C_{10} in Figure 2.6). Let us explain why.

Figure 2.6: Under Wilson's and Grossman's Hypothesis, the Pooling Equilibrium Preferred by the Low-Risk Group Becomes Stable
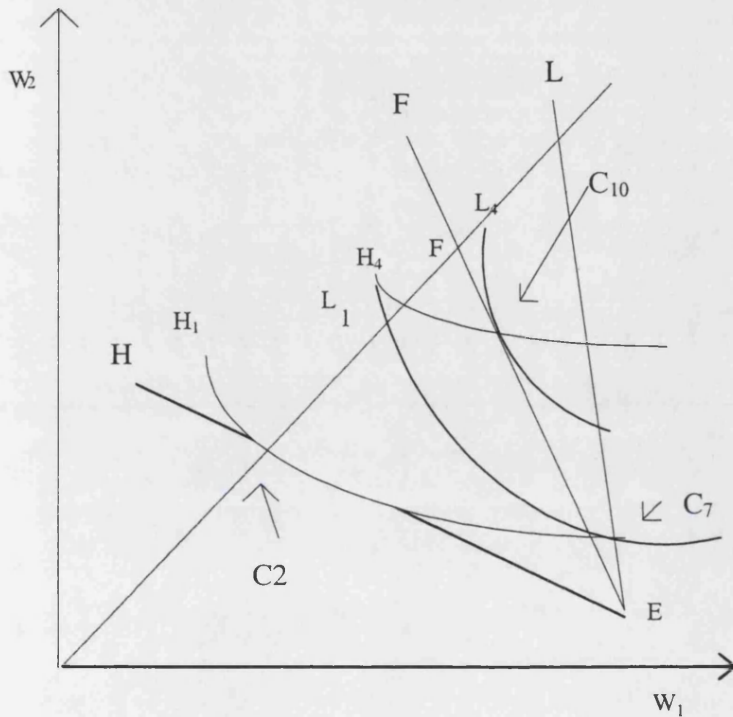
Recall that in the R-S model any new contract offer in the shaded area to the South-East in between L_4 and H_4 was able to upset C_{10} by attracting only low-risk individuals. However, any new contract offer in the shaded area would lead to terminate contract C_{10}, because the latter would remain burdened only by high-risk individuals and it is above the high-risk group zero profit line HE. However, once C_{10} is withdrawn, any new contract in the full area would become unprofitable, because it lies above the zero profit pooling line EF.

Thus, under Wilson's hypothesis, the new contract is not a profitable deviation from C_{10} (see condition III for equilibria discussed above). The same is true under Grossman's hypothesis of strategic behaviour on the part of individuals, because high-risk would immediately 'follow' low-risk individuals in any deviation from C_{10} in the full area, thus making any such deviation immediately

equilibrium in the R-S model.

unprofitable. If they didn't, they would immediately be 'recognized' as being high-risk, and would receive contract C_2 on the high-risk zero profit line EH.

A third contribution to the literature of "equilibrium refinements" is that by Miyazaki and Wilson (M-W, 1977). These two authors adopt Wilson's (1976) definition of equilibrium, but they relax the zero profit constraint hypothesis. According to M-W, loss-making contracts can remain in the market, as long as insurance companies are able to cross-subsidize unprofitable contracts with profitable ones. They show that under this new hypothesis the equilibrium is always characterized by a positive cross-subsidy across contracts, unless the percentage of high-risk λ is greater than a certain threshold λ^{MW}, in which case the R-S separating equilibrium prevails. M-W equilibria tend to become unstable as competition in the market becomes harsher and the incentive to terminate unprofitable contracts increases.

There are several other contributions to the literature on 'equilibrium refinements', which we can only briefly mention in the last part of this section. Riley (1979) extends the R-S's model to a continuum of risk-types under the same behavioural hypotheses, and shows that the insurance market may fail to reach a stable equilibrium under any assumption on the distribution of risk types. Riley also modifies R-S's hypothesis of Nash behaviour on the part of the insurers and develops the concept of Reactive Equilibria, according to which insurers do not offer contracts that they know will become unprofitable if their competitors' reaction is taken into account.

Stiglitz (1978) extends the analysis of insurance market equilibria to the case of monopoly, showing that when there are only two risk groups the only possible equilibrium is separating, in correspondence to which high-risk achieve full insurance and low-risk individuals may be only partially insured. Under monopolistic conditions, low-risk individuals are subject to terms that make them indifferent between buying insurance or be uninsured and, if the percentage of high-risk individuals is large enough, they end up buying no insurance. Thus, although the distribution of surplus is obviously more favourable to the insurer in monopoly than in competition, some of the qualitative results that characterize competitive markets are confirmed and in fact reinforced in the monopoly case.

Newhouse (1996) shows that if there are positive transaction costs, the pooling equilibrium can become stable even under R-S behavioural hypotheses, because the net gain from upsetting any equilibrium contract can be more than compensated by the additional transaction costs necessary to

devise a new contract. Newhouse's intuition is that partial market imperfections in the insurance market may actually be beneficial because they limit the net advantage of underwriting competitors' contractual offers. Ercinosa and Sappington (1997) confirm the same result, by showing that market power and scale economies can facilitate the coincidence between socially preferred and market outcomes in the insurance market (see below). In the spirit of the literature on 'contestable markets', they show that if there are positive sunk-costs (costs of entry that cannot be recovered) there exist market equilibria where the incumbent insurance company cross-subsidizes loss making contracts (those on high-risk) with profitable ones (those on low-risk). The intuition is that if scale economies were significant, any insurance company that served all risk customers would face lower average costs because it would be able to spread its fixed costs over a larger pool of customers. This could more than compensate the potential advantage of screening specific risk types.

The above literature points at the fact that phenomena, which are normally considered negative because they hinder the full display of the "beneficial effects" of competition in other markets, in the insurance market may in fact play a positive role. However, as Newhouse (1984) noted, transaction costs may also exacerbate the effects of the adverse selection problem. If it is difficult to change contract conditions, insurers will be extra-careful before making an offer to a potential high-risk consumer, which may become a permanent liability in the company's budget. If it is impossible or too costly for an insurer to risk-rate a new applicant, the insurer may either reject the applicant or ask a disproportionably high premium.

In summary, the contributions to the literature known as 'equilibrium refinements' highlight that the final characterization of the equilibrium in the health insurance market will depend on the type of prevailing market structure (which determine insurance companies' conduct and constraints). Some of R-S's original conclusions are confirmed, and others are further specified, such as those concerning the stability of pooling equilibria and of equilibria characterized by positive cross-subsidies, under different hypotheses concerning agents' behaviour and market imperfections.

2.7 Second Best (SB) Equilibria

A second stream of literature set the foundation for welfare analysis under conditions of imperfect information, by looking at the properties of Second Best equilibria. Such equilibria emerge when the

individuals' maximization problem is not only constrained by resource availability but also by agents' incentive-compatibility constraints.

The core concept, first developed by Harris and Townsend (1981) and then explicitly applied to the insurance market by Crocker and Snow (1985), is that of constrained Pareto-efficiency. A market allocation is constrained-Pareto efficient, or second best, if it is Pareto efficient among all possible allocations satisfying:

- resource constraints;

- incentive compatibility, or self-selection constraints.

In other words, Pareto-constrained are 'the best' equilibria (in terms of agents' utility) that can be achieved whenever a self-selection or incentive compatibility constraint (according to which agents must be induced to reveal their private information; see Chapter 5) is considered alongside a resource constraint. Let us define $C^h = \{\alpha^h_1, \alpha^h_2\}$ the set of contracts meant for the high-risk group and $C^l = \{\alpha^l_1, \alpha^l_2\}$ the set of contracts meant for the low-risk group. The formal structure of the problem to characterize SB equilibria is the following:

$$\text{Max}_{\alpha l\, \alpha h}\ \mu V(p^h, C^h) + (1-\mu)V(p^l, C^l) \qquad (2.9)$$

Subject to:

1. $\lambda\pi(p^h, C^h) + (1-\lambda)\,\pi(p^l, C^l) = \lambda[(1-p^h)\,\alpha^h_1 - p^h\alpha^h_2] + (1-\lambda)[(1-p^l)\,\alpha^l_1 - p^l\alpha^l_2] \geq 0$

   (Resource, R, constraint).

2. $V(p^h, C^h) \geq V(p^h, C^l)$ and $V(p^l, C^l) \geq V(p^l, C^h)$

   (Incentive-Compatibility, IC, constraints).

μ is an arbitrary weight given to high-risk in the welfare function: $\mu \in [0,1]$.

SB contracts are Pareto-efficient contracts among those satisfying constraints 1 and 2 above. If we exclude equilibria characterized by over-insurance [33] (where individuals would end up with higher wealth in case the 'accident' occurs), it is possible to prove that all SB equilibria (see Harris and Townsend, 1981) satisfy the following necessary conditions:

a) C^h provides full insurance to high-risk.

---

[33] Formally, we are imposing that $\lambda \geq \mu$, i.e., the proportion of high-risks is higher than their weight in the social welfare function.

b) High-risk are indifferent between C^h and C^l

c) Any losses (gains) on high-risk contracts C^h are exactly compensated by gains (losses) on low-risk contracts C^l.

In Figure 2.7, the segment C_7F (a subset of the curve J^hJ^l) identifies the set of (separating) contracts for low-risk that, in combination with full insurance contracts for high-risk individuals along C_2F, satisfy conditions a), b) and c) above, and are thus potential candidates for SB equilibria. Note that C_2-C_7, the R-S separating market equilibrium discussed before, belongs to the set of potential SB equilibria candidates. In correspondence to C_2-C_7 both contracts break even, and there is no cross-subsidy between risk groups: condition c) above is trivially satisfied.

Figure 2.7: Second Best Equilibria
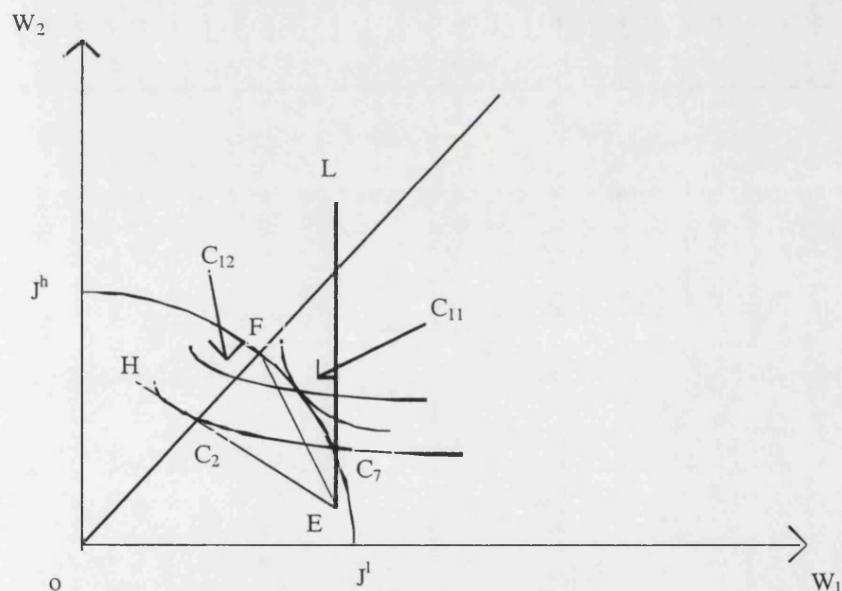
Let us identify the second best contracts among all the possible candidates (contracts that satisfy conditions a, b and c above) in a graphical example, illustrated in Figure 2.7. As we move along J^lF starting from C_7 and correspondingly move towards F from C_2, the cross-subsidy from low risk in favour of high-risk becomes positive: in correspondence to these contracts, low-risk are paying progressively more than a "fair-odds" premium (moving to the left of segment EL along J^lF) to receive more complete coverage (moving towards the bisetrix); by contrast, high-risk individuals are paying progressively less than a 'fair-odds' premium and still obtain full coverage (moving to the right, or North-East, of segment EH along the bisetrix). Note that in Figure 2.7 by moving from C_7

along J^lF low risk individuals initially increase their utility level (graphically, they are able to reach progressively higher indifference curves), even if they are paying more than a fair-odds premium for this additional insurance. They are able to receive more complete coverage than in C_7, while still separating from high-risk. Their preferred equilibrium is in correspondence to point C_{11}, at the tangency of J^hJ^l and their highest indifference curve. In parallel, high-risks increase their utility moving from C_2 to C_{12} (they still receive full-insurance and pay progressively less for it). Thus in Figure 2.7, C_2-C_7 is not a SB equilibrium[34], and by introducing the possibility of a positive cross-subsidy across risk-groups, one can move from Pareto-dominated towards Pareto-superior equilibria (such as C_{11}-C_{12}).

If we further increase the degree of redistribution or cross-subsidy across risks beyond C_{11}-C_{12} (converging towards F from C_{11}-C_{12}), high-risks' utility further increases, while low-risks' utility decreases. Thus, in is not possible to increase both groups' utility beyond C_{11}-C_{12} and still respect necessary conditions a,b and c for SB equilibria. All combinations of contracts along FJ^l from C_{11} to F for low-risk, which correspond to contracts from C_{12} to F for high-risk, are not comparable according to the Pareto criterion, and they all represent (constrained) Pareto-optimal or SB contracts. F is the pooling SB contract preferred by high risk individuals. In correspondence of F the cross-subsidy is highest.

Note that the above analysis of SB equilibria provides a significant, yet unusual, rationale for redistribution across risk groups in the health insurance market. Traditionally, the desirability of cross-subsidy from low to high risk groups has been justified on the basis of equity, by pointing out that it ought to be a public responsibility to care for the health needs of those, like the old and the chronically ill, who would otherwise be unable to buy insurance at a premium that reflects their health risk, or to pay directly for services. Or it has been justified by appealing to a life-cycle argument: the young and healthy (low-risk) accept to pay more because they know that in the future, when they will become old and sick, they will in turn be able to benefit from subsidized health insurance coverage (see Annex 2.1). The above analysis provides a third justification of redistribution, by showing that even without considering the future, in fact it may be welfare improving and in low-risk individuals' interest to provide some form of subsidy in favour of the high-risk, as a means to achieve a more complete level of health insurance coverage for themselves.

---

[34] Instead, see Figure 2.8 for a case where C_2-C_7 is a second-best equilibrium.

2.8 Optimality of Market Equilibria

It is possible to utilise the conceptual framework sketched above as a benchmark to analyse the optimality of private insurance market equilibria and the welfare impact of different government interventions in the health insurance market, according to different behavioural hypotheses. As we showed in the preceding sections, if in equilibrium each contract must break even, the market cannot provide both separation of risks and cross-subsidization. Thus, if a positive level of cross-subsidization characterizes SB equilibria, the R-S separating equilibrium is not SB optimal (in Figure 2.7, we indicated that C_2-C_7 the R-S separating equilibrium is Pareto-dominated by C_{11} C_{12}). However, this last conclusion does not necessarily hold, and we now turn to a graphical example where it doesn't, illustrated in Figure 2.8.

Figure 2.8: Case Where the R-S Separating Equilibrium Is SB Optimal
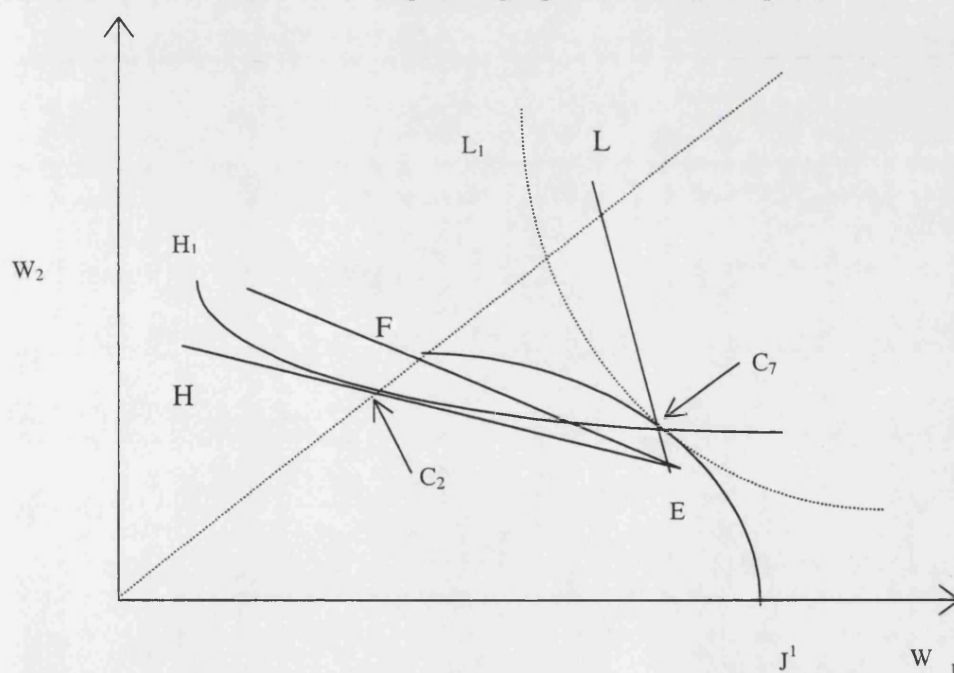
In the above figure, C_7 is at the point of tangency of J^1F and the low-risk group's map of indifference curves (in correspondence to L_1). Thus, any further movement towards F along J^1F would lead to a decrease of the low-risk's utility level, and would not be a Pareto improvement.

In conclusion, the (Second Best) optimality of market equilibria depends on the value assumed by some key parameters (among others, individuals' degree of risk-aversion, the probability of the event which individuals buy insurance against, and the variability and volatility of the risk in the insured population). Which characterization of these parameters leads to a situation such as that described in Figure, 2.8, as opposed to that prevailing in Figure 2.7? The literature indicates that the R-S separating equilibrium is more likely to be SB optimal, as it is in Figure 2.8, when the percentage of high-risk individuals is sufficiently large, or when the degree of risk aversion of low risks is low, or when the risk differential among different risk groups is higher. The intuition is that all of these parameters increase the cost of cross-subsidizing high-risk individuals. For example, if the proportion of high risk individuals is above a certain threshold λ^{MW}[35], the extra benefit low-risk gain from increasing their insurance coverage relative to the R-S equilibrium level is not sufficient to justify the extra costs they must sustain to cross-subsidize all high-risk individuals. The same result holds if low-risk individuals are not very risk-averse, or if their probability of incurring the event for which they could buy insurance coverage is low. It turns out that the threshold λ^{MW} is greater than λ^{RS}, the percentage of high-risks beyond which the R-S separating equilibrium exists and is stable, under Wilson's as well as Grossman's hypotheses. This leads us to state the following important result: there is an interval: $\lambda^{RS} < \lambda < \lambda^{MW}$, where λ describes the share of high-risk individuals over the total, in correspondence to which the R-S separating equilibrium exists, but it is (Second Best) sub-optimal.

By contrast, when $\lambda < \lambda^{RS}$, under Wilson's as well as Grossman's behavioural hypotheses, the pooling equilibrium preferred by low-risk is stable. Such pooling equilibrium is never SB optimal because it violates condition a) above, valid for all SB equilibria, i.e., it does not provide full insurance to high-risk.

However, the pooling equilibrium may not be Pareto comparable with the second best pair of equilibrium contracts preferred by low-risks[36], or with the second best equilibrium contract preferred by high-risk[37]. The comparison between pooling equilibria and SB contracts is illustrated in Figure 2.9 below. In general, from a pooling equilibrium it is possible to reach Pareto-superior points by

---

[35] It turns out that the Miyazaki-Wilson pooling equilibrium exactly coincides with the second best equilibrium preferred by low-risks whenever $\lambda < \lambda^{MW}$. Otherwise, under the M-W hypothesis, the R-S separating equilibrium is selected. Thus, M-W and SB equilibria coincide.

[36] Contracts C_{11} and C_{12} in Figure 2.7 may be characterized by a level of utility for high-risk lower than in the pooling equilibrium.

[37] Point F in Figure 2.7 may entail a level of utility for low-risk lower than in the pooling equilibrium

allowing a restructuring of cross-subsidies that separates the different risk groups (moving towards the frontier in a North- East direction in the figure).

Figure 2.9: Second Best and Pooling Equilibria

F=Full insurance pooling
equlibrium

Set of 2nd best Pareto optimal
equilibria

Ut High Risks

Market
pooling
equilibrium

$C_{11}C_{12}$=Second best
equilibrium preferred by
low risks

Ut Low Risks
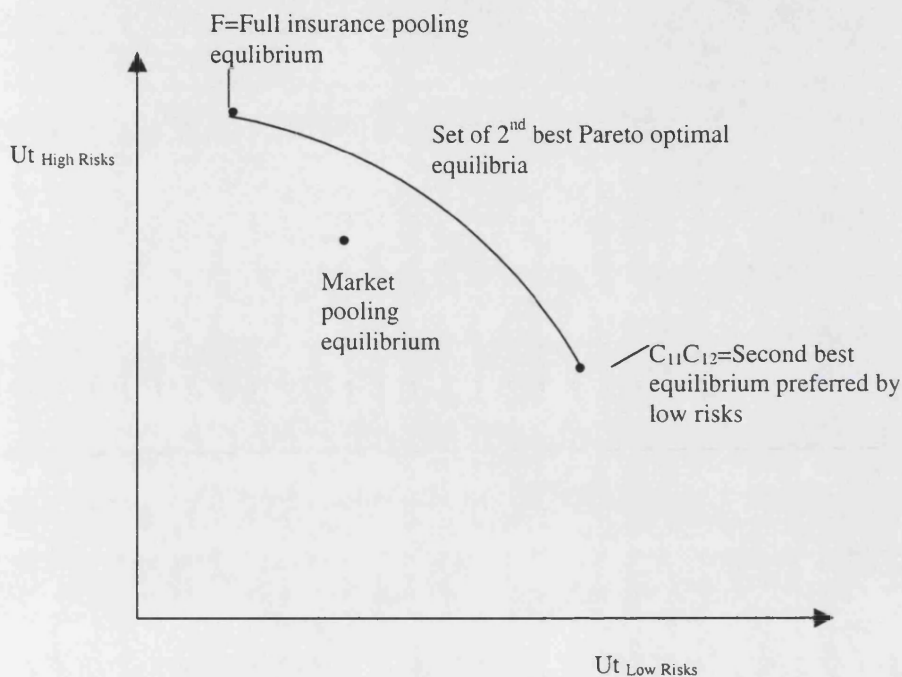
2.9 Policy Analysis

In this and the following sections we shall analyse the effects of different policy options meant to correct health insurance market equilibria, using the analytical framework first developed by Neudeck and Podczeck, 1996 (hereafter N-P). The set of policy options we consider is the following:

1. Public provision of insurance/subsidies.

   1.a Full public insurance.

   1.b Partial compulsory public insurance, without or 1.c with the possibility of acquiring supplementary coverage from the private sector (the so-called 'topping up' of insurance).

   1.d Full public insurance with the possibility of opting out.

1.e Risk-adjusted premium subsidies.

2   Regulation of the private insurance market.

   2.a Standard contract with full-coverage.

   2.b Minimum insurance.

   2.c Premium rate restrictions.

In discussing all the above 8 cases, we will maintain Grossman's hypothesis concerning dissembling behaviour on the part of high-risk individuals (see section 2.6). According to this hypothesis, no loss-making and cross subsidized contracts are maintained in the market equilibrium.

2.9.1 Public Provision of Insurance

Let us first analyse the options of full or partial public insurance. In the latter case the government may or may not allow individuals to purchase supplementary insurance from the private sector. Let us suppose that under public insurance all risk groups are required to contribute the same amount to the scheme (thus forcing a cross-subsidy from low to high risk individuals), and the public insurance scheme must on average break even (the expected total amount of contributions must equal the expected total benefits disbursed).

Figure 2.10: Full Public Insurance and Partial Public Insurance
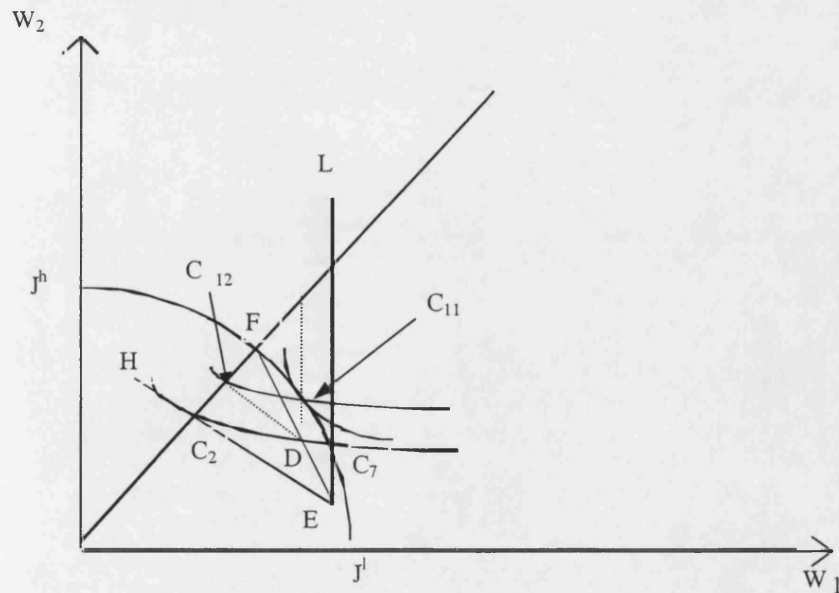
Graphically, one can see that starting from the initial endowment E the introduction of a compulsory full insurance scheme would be equivalent to 'forcing' both risk groups along the pooling line EF up to the pooling, full insurance equilibrium in F, which is the SB equilibrium preferred by high-risk individuals. Point F is characterized by the greatest level of redistribution across risk types and may be not Pareto-comparable with the market separating equilibrium C_2-C_7, because low risk may be better off in C_7 than in F.

Partial public insurance brings both risk groups along the pooling line EF towards, but not 'all the way to' the bisetrix, say to point D. Figure 2.10 shows that, starting from point D, it is possible to reach Pareto-superior points by allowing individuals to purchase supplementary insurance from the private sector. By these means, in the above figure from point D the separating contracts C_{11} -C_{12} can be reached. Contract C_{11} is the preferred one by low risk individuals among second best contracts.

Thus, the above analysis highlights two important results: 1) moving from a situation of public full insurance to one of partial insurance is likely to lead to a decrease in the high-risks' utility. 2) From a situation of partial public insurance, it is always optimal to allow a 'top-up' of public benefits for all risk-types. In fact, by varying the "generosity" of the public insurance scheme (each level of coverage corresponds graphically to a point along segment EF) and by allowing supplementary private insurance all second best contracts can be obtained as separating market equilibria. The more comprehensive is the extent of public insurance coverage, the greater is redistribution from low to high-risk individuals and the higher is the utility enjoyed in equilibrium by the latter group.

The third type of corrective intervention in the health insurance market we consider is the provision of a full public insurance scheme, with the possibility of opting out. In situations characterised by universal compulsory public insurance schemes, those who believe in competition as a means to stimulate greater efficiency in the health care market are in favour of this possibility (for the experience in Chile, see following Case study). The claim is that allowing those who are dissatisfied with the public scheme to opt out and to enrol with private insurers would create incentives for greater efficiency and better quality services. Instead, those who oppose such move towards liberalization of the health insurance market claim that private insurers would just cream-skim the "good" risks and leave all the "bad" risks to the public scheme. In the language of the R-S model, the public scheme would lose the low risk individuals' and would remain burdened with all the high-risk individuals.

Let us illustrate this case in Figure 2.11.

Figure 2.11: Full Public Insurance with Possibility of Opting Out
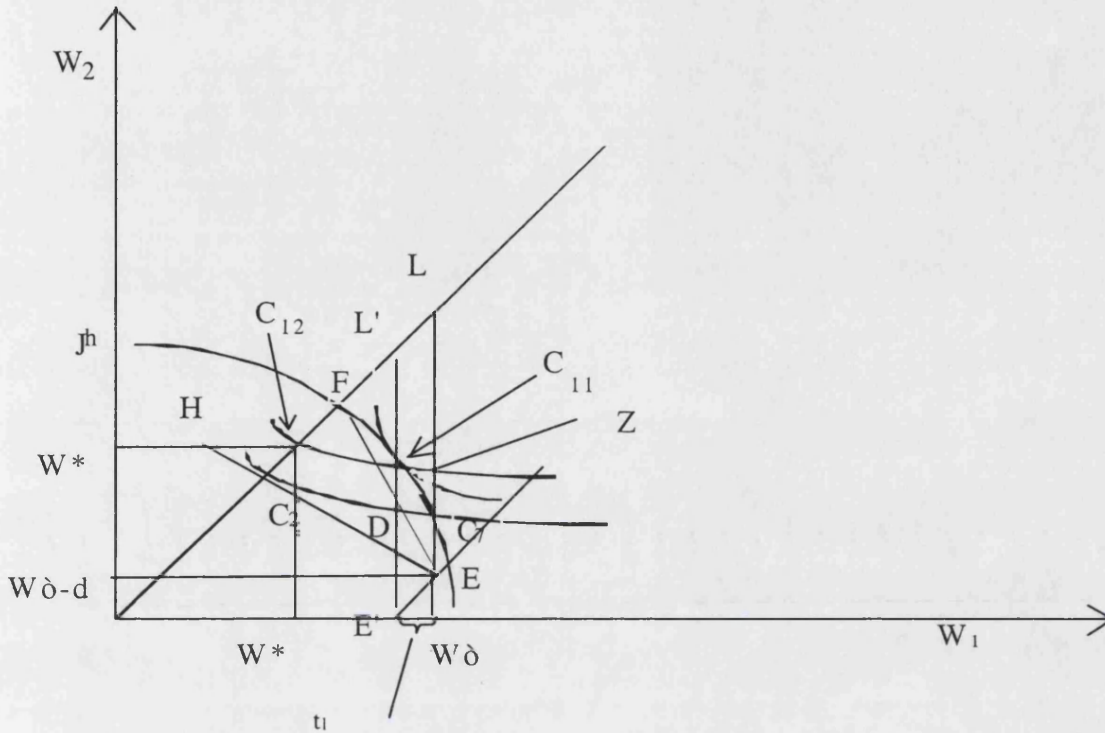
Suppose that starting from E the government initially offers insurance contract C_{12} to everybody. The contract is characterized by full insurance and it is preferred to the separating contracts C_2C_7 that would prevail as the competitive equilibrium only by high-risk individuals. Note that if purchased by everybody the contract would be profitable (it is below EF, the pooling zero profit or 'fair-odds' line), but if purchased only by high-risk individuals it needs to be subsidized, because it is above the 'fair odds' line for high risks, EH. Suppose that a (net) proportional tax t is imposed on all individuals who decide not to purchase C_{12}.

If purchased only by high-risk, the total per capita loss on contract C_{12} would be equal to W*-Wò+t+p^hd (contract C_{12} would guarantee with certainty wealth W* instead of an expected after-tax endowment equal to Wò-t-p^hd). To compute the total subsidy it is necessary to determine the fraction of net beneficiaries to taxpayers, which, under the hypothesis that C_{12} is purchased only by high risk individuals, in equilibrium is equal to the fraction of high-risk individuals in the population. As before, suppose that the population fraction that is high risk is equal to λ. Then, the lump sum (capitation) tax on initial endowments necessary to sustain contract C_{12} would be equal to:

$$\lambda t_1 + (1-\lambda)t_1 = \lambda(W^* - W\delta + t_1 + p^h d)$$

$$t_1 = \lambda/(1-\lambda)(W^* - W\delta + p^h d) \hspace{4cm} (2.10)$$

In Figure 2.11, the tax switches the low risk individuals' initial endowment from E to E' and their fair odds line from segment EL to E'L'[38]. After the tax, low risk individuals (unlike high-risks) would still be willing to opt out of C_{12}, and purchase a new contract such as C_{11}.[39] Contract C_{11} is actually the best insurance contract low risk can receive along their after tax 'fair odds' line (E'L') among all the contracts high risk individuals would not prefer vis-à-vis C_{12} (so it is the best contract compatible with separation between risk groups). Finally, note that, being on the after-tax 'fair odds' line (E'L'), a private insurer would break even by offering C_{11}, and by attracting only low-risk individuals[40]. Thus, C_{11}-C_{12} is a candidate SB equilibrium, and it is possible to show that it is in fact the only feasible SB equilibrium, given the government's full insurance contract offer in C_{12}. C_{11}-C_{12} is in fact the same SB equilibrium preferred by low risk individuals we analysed in Figure 2.7. It is possible to prove that starting from a different initial full insurance contract on segment C_{12}-F and thus from a different tax on initial endowments, any of the equilibria that belong to the set of SB equilibria can be reached.

We can utilize the same conceptual framework, to illustrate the effects of a risk-related premium subsidy in favour of high-risk individuals, funded through a mandatory contribution to a solidarity fund imposed on low risks. The subsidy would be earmarked to purchase a specified health insurance plan (in Figure 2.11, the full coverage plan C_{12}), and it would be exclusively based on individuals' relevant risk characteristics (the lower the risk, the higher the contribution, or the lower the subsidy). Unlike a premium tax deduction, it would be unrelated to the premium amount that individuals pay[41]. The subsidy would increase the affordability of the full-insurance contract for the high-risks (who, in Figure 2.11 can purchase contract C_{12}, above their 'fair-odds' line), and at the same time it would

---

[38] A similar shift, for clarity not represented in the above figure, would occur to high-risk individuals' budget line.

[39] Note there is no contract offer along the zero profit pooling line EF that low-risk individuals would prefer vis-à-vis C_{11} (so there cannot be any pooling equilibrium).

[40] Any attempt to further undercut C_{11} in order to attract low risk individuals (for example, with a new contract offer in Z, or any other point in the shaded area in-between the two indifference curves passing through C_{11}) would lie to the right of E'L', and would thus be unprofitable.

[41] When the amount of the deduction increases with the total premium that individuals pay, their "fair odds line" (which is their budget constraint, showing the best rate at which they are able to exchange wealth between the two states) changes inclination. Then we would observe a substitution as well as an income effect, and most likely a more comprehensive insurance coverage in equilibrium (unless a large negative income effect compensated the substitution effect).

shift the low-risks' 'fair-odds' line from EL to E'L' [42]. By paying the premium subsidy to the high-risk, in fact low-risk individuals would be able to achieve higher insurance coverage.

Equal results would also be obtained by imposing an opting-out fee to be paid only by those (low-risk individuals in our example) who deviate from the initial full-insurance contract. This would be the case if government levied on all individuals a tax equal to t_1, and at the same time opened up the possibility of opting out of mandatory insurance (contract C_{12} in Figure 2.11)

Drawing a conclusion from all of the above findings, we can say that, although risk selection does occur when the insurance market is liberalized (the pooling equilibrium is upset, and a separating equilibrium prevails), the government may still achieve second best outcomes, provided low risks can be forced to continue to contribute to the public scheme when they are not any more benefiting from it. In the case of Chile, discussed in Case Study 2.1, that is precisely what seems to have happened: the wealthier segments of the population have in theory been allowed to 'opt-out' of the public scheme (FONASA) and enrol in private schemes (ISAPRES), but in practice they have been required to continue to contribute to the public scheme. High (low) risk individuals are made (better) worse off than in the full public insurance pooling equilibrium after the introduction of the possibility of opting out; yet, SB equilibria can still resist if a compulsory cross-subsidy across risk groups is maintained[43].

Van de Ven et al. (2000) and Van de Ven and Ellis (2000) provide a detailed analysis of the different ways through which risk-adjusted cross subsidies may be implemented, and of the different possible risk-sharing mechanisms among different clients, insurers and a solidarity/compensatory fund (see also Chapter 6). In the implementation of the risk-adjusted subsidies two main issues arise. The first concerns the criteria to define the different risk categories, and the second the criteria to compute the subsidy/contribution for the different risk categories. In the above example, with just two risk groups and only one subsidised "residual" public insurance contract, the low risk self-selected themselves by opting out of the public scheme. With more than two risk categories and several plans, van de Ven suggests adopting a nationwide, standard rating model as a basis to compute the subsidy value per risk category. He also proposes that the government mandates that insurance

---

[42] As before, insurers' interest to selectively attract low risks through contract in the shaded area (with contracts such as Z) would be eliminated.

[43] Another critical issue, not considered above, concerns the difficulty of maintaining a political consensus for the cross-subsidy in favour of the high risks, once it does not provide any more direct benefit to those who are supposed to remain its net-contributors. These issues are similar to those discussed by Hirshman (1970) in assessing the relative strengths and weaknesses of "voice" vis-à-vis "exit" mechanisms (see Chapter 6, and Annex 6.5 on the political limits to redistribution.

companies share their information concerning their clients' health "riskiness". He claims that by these means insurers' incentive to select risks would be contained, leaving open the other beneficial dimensions of competition (on quality, cost, etc.). For the computation of the cross-subsidies, Newhouse (1989) proposes that a combination of prospective risk-adjustment methods, and of retrospective health cost/utilization-information could be used. Purely prospective risk-adjusters are able to capture only a small fraction of the variability in individuals' future health expenditure, while just using past health expenditure would hinder insurers' incentives to improve efficiency, and to search for the more cost-effective providers in the market.

Case Study 2.1: Health Insurance Reforms in Chile[44]

Initiated by the Pinochet regime in the '80s, health reforms in Chile allowed people in the higher income groups to enrol with private insurers (ISAPRES). The government continued to directly finance a sub-set of providers, used by the non-insured population, or by those insured with the public residual insurer (FONASA). Health reforms in Chile were severely criticized, on the basis of arguments similar to those we discussed in this Chapter. The critics claimed that, in the absence of a regulatory framework, private insurers would cream-skim the good risks (the rich and healthy), whilst the public sector would remain burdened with the poorer and less healthy segments of the population.

Following the reforms, the predicted segmentation of the health insurance market into those privately insured (approximately one fourth of the population) and those left with public coverage did occur. FONASA, the "residual" compulsory public insurance scheme, currently covers virtually the entire elderly, chronically sick and economically weak segments of the population. However, the evidence also shows that such evolution has not been accompanied by a tangible worsening of health utilization or health indicators for the poor. Two properties still characterised the Chilean health financing system after the reforms:

1. The government still required those who decided to opt out of the public system and who enrolled with ISAPRES, the private schemes, to contribute through taxation to the public scheme FONASA. [45] Over the '80s and '90s, the public scheme actually received significant additional funding (see below).

2. Unlike several other Latin American Countries, the public subsidy in favour of the private insurance schemes ISAPRES, which serve the upper segments of the population, has been kept at negligible levels. The government only spends 4% of its general revenues to subsidize health expenditure of

---

[44] Most of the evidence hereafter reported is from W. Savedoff, 2000, Is Anybody Listening? Ignoring Evidence in the Latin American Health Reform Debates.

[45] Moreover, contributions were set as a fixed percentage of income with no capping, so that there can be hardly any price competition among insurance companies to attract healthier customers by underwriting

those who opt out of the public insurance scheme (all of which are employed and with incomes above the average).

Milanovic (1995) compared distribution of benefits from the public health system in Chile and in the UK and Hungary, which, unlike Chile, have preserved a universal public insurance system, and concluded that Chile is characterised by the most progressive distribution of benefits.

Figure 2.12: Distribution of Public Sector Benefits in Health Services.
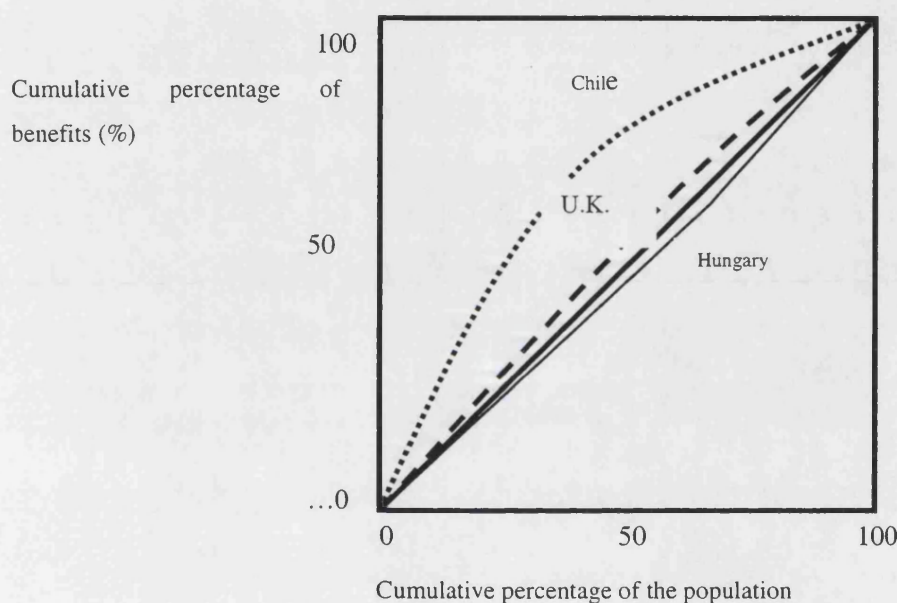
Source: Savedoff (2000), based on Milanovic (1995)

Bitrán (1998) confirms the same results: in Chile 72% of the public health system beneficiaries are from income groups earning less than US$144 per month. The above results are not surprising, given that in Chile, unlike Hungary and the UK, public facilities are mainly used by those who are enrolled with FONASA, who belong to the poorer segments of society.

More importantly, the evidence available shows that the health benefits offered by FONASA are currently quite comprehensive, and that the health system continues to be easily accessible to all. Sapelli and Vial (1998), using need-adjusted health service utilization measures, show that the distribution of utilization is rather even across income groups, and that the poor do not seem to be overly discriminated by the existing two-tier system. By looking at the distribution of medical visits and days of hospitalisation by socio-economic

---

existing contracts.

58

group (see Figure below), they conclude that there is no significant relationship between income and utilization.

Figure 2.13: Expenditures on Services by Income Quintile
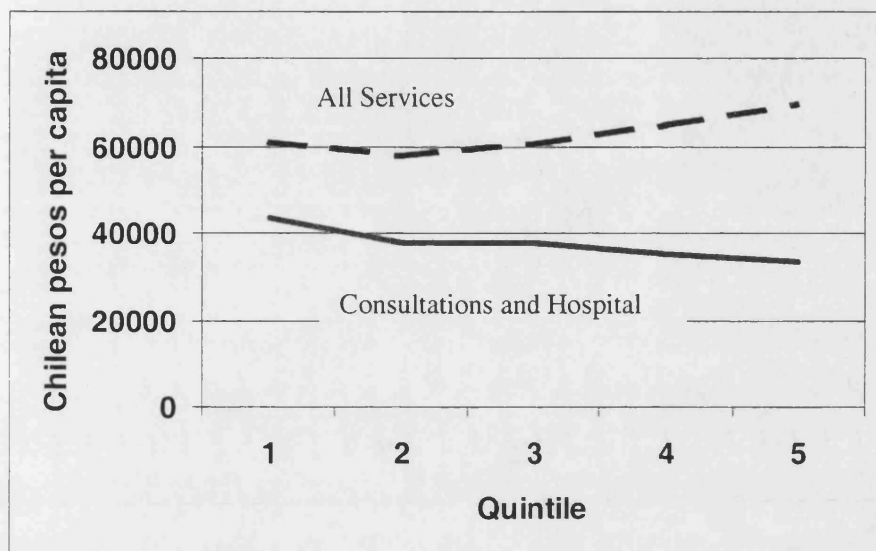
2.9.2 Regulation of the Private Insurance Market

In alternative to direct public provision/subsidization of insurance, the government could try to regulate the private insurance market. This is the alternative advocated by those who consider private insurers more effective than public agencies at managing health insurance funds, and at stimulating greater efficiency, and quality of services from providers. According to this view, governments should not impose any cross-subsidies, but simply impose a regulation on the market so that all can enjoy a minimal level of benefits. N-P (1996) show that this argument is actually mistaken, by proving that whichever form of regulation is imposed on the market (unless it is regulation forcing subsidies across risk-types, and limiting insurers' ability to select clients), it will not eliminate the possibility of chronic instability and, more importantly, the incentive to risk-select[46].

---

[46] This section also reviews and builds upon N-P's (1996) analysis.

The first form of regulation considered consists of an obligation imposed on all private insurance companies to offer a standard contract with full coverage, open to any individual who wants to purchase it. In this case private companies would be free to offer any other contract they wish, alongside the standard contract. N-P show that with this form of regulation no equilibrium may exist. Let us consider Figure 2.14 below.

Figure 2.14: Standard Contract With Full Coverage
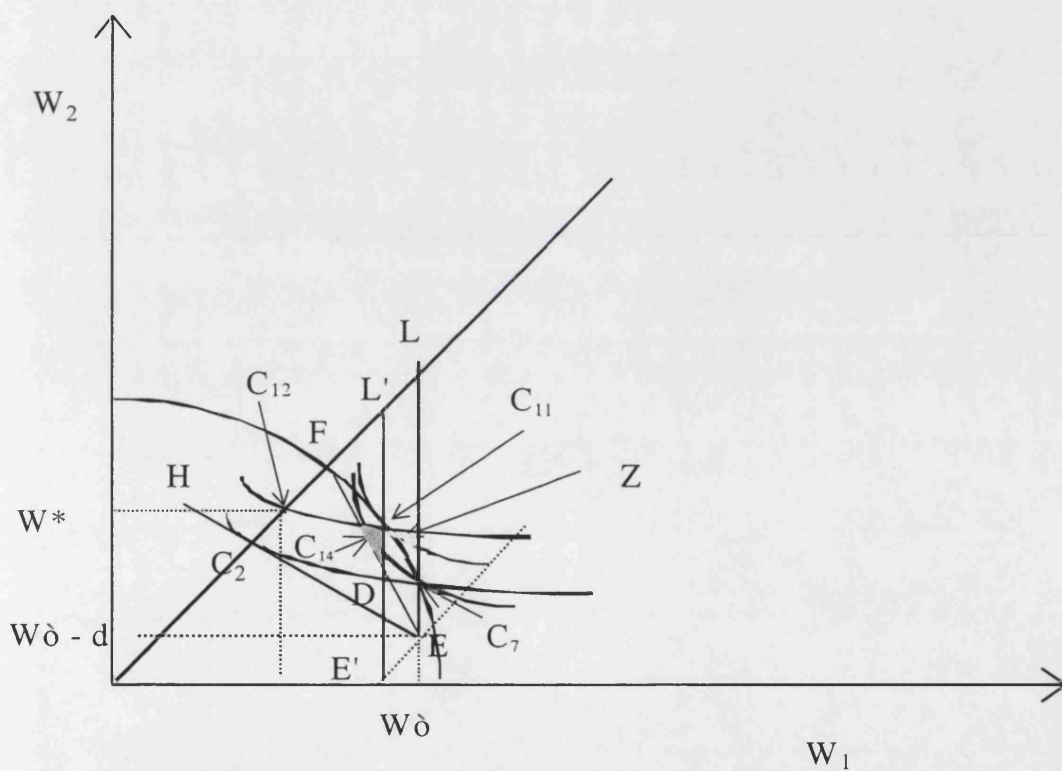
Suppose that the initial market equilibrium is the R-S separating equilibrium C_2-C_7 described in section 2.5. The standard full insurance contract imposed by regulation must lie on segment C_2-F

along the bisetrix (contract offers to the south-west of C_2 are ineffective, as nobody would choose them, while contract offers to the north-east of F are unfeasible). Assume that such standard contract is in C_{12}. As we showed before, low-risk individuals have always an incentive to deviate from C_{12} if other contracts are offered in the market. For example, consider contract C_{14}, in the shaded area above (below) the low (high) risk indifference curve through C_{12} and to the left of the low-risk after tax fair odds line. If C_{14} were offered, low risk would accept it and stop purchasing the standard contract C_{12}. Then, no pooling equilibrium exists, and in order to sustain C_{12} as part of a separating equilibrium, it is necessary to maintain a positive cross-subsidy from low to high-risk individuals (see discussion in section 2.9.1, and Figure 2.11). However, market competition to attract low-risks will make such cross-subsidy unsustainable. From C_{14} competition to attract low risks will lead to further lower (increase) the premium (indemnity) charged (given) to them. With a risk-related subsidy, we showed that the equilibrium contract for low-risks would be found on C_{11}, along segment E'L' [47]. E'L' was the budget line for low risks when they were forced to subsidize the standard contract C_{12}, and it would still be, were all insurance companies enrolling the same share of high and low risk individuals. However, insurers can try to further undercut C_{11} by new offers such as Z, which are profitable (being to the left of EL) as long as the insurer who is offering them is able to attract all low risk and to maintain a share of high-risk individuals equal to that of its competitors. Thus, the standard contract policy leads to chronic market instability, unless all insurance companies are required to have the same proportion of high and low risk clients [48].

Let us clarify illustrate this point by a numerical example. Consider a situation where each high-risk individual receives \$20 over his/her "fair odds" premium in the standard contract C_{12}. Suppose there are 10 high-risk individuals, so that the total cross-subsidy to high risk is equal to \$200. Suppose there are two competing insurers in the market, each of them shouldering half of the high-risk individuals, so that the total cross-subsidy that each insurance company needs to collect from low risk clients to sustain the standard contract is equal to \$100. If there are 20 low risk individuals, and if low risk clients are evenly distributed between the two competing insurers, each of them would have to pay \$10 = \$20 (1/3)/(2/3)[49] above their fair-odds premium to subsidize the standard contract for high risk. In terms of figure 2.14, such cross-subsidy would shift the low-risk clients' zero profit line to E'L'. However, suppose that one of the two insurance companies offered a new contract,

---

[47] C_{11}-C_{12} is the only possible separating equilibrium when E'L' is the low risk individuals' fair-odds line. See Figure 2.11 and relative explanation in the text.
[48] As C_{12}, the standard contract, must remain on offer, C_{11}-C_{12} would be the only possible separating equilibrium if all insurance companies shared the same risk composition of clients.
[49] See expression (2.10) in the text.

such as Z in Figure 2.14, which attracts all low-risk individuals, at the same time maintaining an equal share of high-risk individuals (who are still purchasing the mandatory standard contract C_{12}). In the new situation, the amount that each of the 20 low risk individuals that shifts to contract Z should have to pay above their fair-odds premium to cross-subsidize the standard contract would be only \$5 [\$5 = \$20(1/5)/(4/5)]. At the same time, the other insurer, burdened with half (5) of the high-risk individuals, and with no low-risk individuals left, becomes unable to subsidize the standard contract. Thus, the imposition of a standard contract may cause chronic instability in the market, as competitors try to undercut each other to attract low risk individuals.

The second form of regulation we consider consists of a minimum insurance requirement. In the graphical representation below, a minimum insurance requirement is considered equivalent to imposing a lower bound on the wealth level in case the accident occurs, measured on the vertical axis.

Figure 2.15: Imposition of a Minimum Insurance Requirement
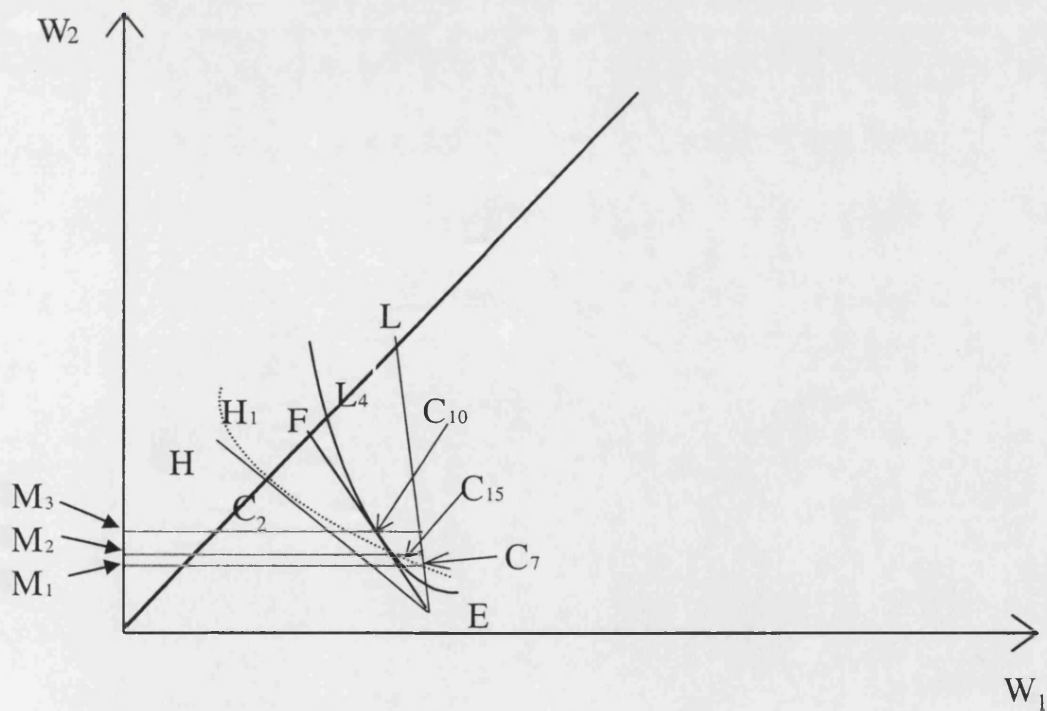
In Figure 2.15, if we initially consider a situation where the R-S separating market equilibrium $C_2 C_7$ prevails, we can distinguish among three cases:

- A minimum insurance requirement comprised between M_1 and M_2. In this case the minimum insurance constraint is binding only for low risk individuals, while high-risk can continue to purchase C_2 in equilibrium. The high-risk individuals' indifference curve through C_2 (H_1 in figure 2.15) lies above the highest indifference curve for low risk along the pooling line EF (L_4 in Figure 2.15), which implies that the utility low risk individuals can gain by still separating from high risk (choosing a point along H_1 between C_7 and C_{15}) is higher than the maximum utility they can achieve in any pooling equilibrium. Then, low risks choose a separating contract along H_1, between C_7 and C_{15}, while high risk will choose the separating equilibrium contract C_2.

- A minimum insurance requirement between M_2 and M_3. In this case low risk individuals prefer to switch to the pooling equilibrium C_{10}, which gives them a higher utility (L_4 in Figure 2.15). The high risk group will also switch to contract C_{10}, which provides them with a utility level higher than C_2. This pooling equilibrium is stable, because any deviation to the "right" of the pooling line EF, meant to selectively attract low risk, will in fact also attract high-risk individuals, and will then be unprofitable.

- A minimum insurance requirement at a level above M_3. In this case the equilibrium will be found on the pooling line EF in correspondence to the intersection with the minimum insurance requirement. In the extreme case, the regulator can impose the full insurance pooling equilibrium in F.

Hence, setting the minimum insurance requirement in correspondence to more comprehensive levels of coverage progressively forces the pooling equilibrium in the market. Note that in general the equilibrium with a minimum insurance requirement is not second best and cannot be compared with the market equilibrium. However, by imposing a minimum insurance requirement the regulator can limit insurers' practice to undercut existing contracts in order to attract low risk, and can stimulate a positive cross-subsidization across risk types.

The last form of regulation we consider is premium rate restrictions[50]. In terms of the above graphs, such form of regulation entails a restriction on the allowable gap between the rates charged to high-risk and to low-risk individuals, which is equivalent to forcing insurers to maintain a set of loss

---

[50] To prevent insurers from rejecting new applicants or from cream-skimming high risk individuals, the regulator may complement premium rate restrictions with a periodic open enrolment requirement.

making contracts that need to be subsidised. However, similarly to the case of the standard contract regulation, imposing rate-restrictions may actually exacerbate the impact of adverse selection, as insurance companies struggle to dilute the share of loss-making contracts in their pool by discouraging subscription of high-risk individuals, and by fiercely competing for low risk individuals.

2.10 Conclusion

This chapter has discussed the consequences of adverse selection on the functioning of insurance markets. To isolate the effects of adverse selection from other confounding factors, it has considered a benchmark situation with no moral hazard and characterized by perfect competition. We have shown that the full information equilibrium is characterized by complete insurance coverage, and by different rates charged to the low and high-risk individuals. However, with incomplete information insurance companies compete to attract low risks, by offering less expensive contracts characterized by less comprehensive coverage. The separating equilibrium is characterized by less than optimal coverage, and, if insurers and insurees behave myopically, no stable equilibrium may exist. By contrast, second best equilibria are generally characterized by a positive cross-subsidy across risk types.

The government can intervene in the health insurance market in two ways: by directly providing/subsidizing insurance, or by regulation. Following Neudeck and Podczeck (1996), we show that the two forms of intervention do not lead to identical results.

Provision/subsidization of insurance, supplemented with private insurance or with the possibility of opting out, can lead to second best equilibria. When they are characterized by positive cross-subsidies between risk-types, SB equilibria demand that the government is able to subsidize contracts with higher than average premium/benefit ratios, and to tax contracts with lower than average premium/benefit ratios.

By contrast, regulation of the private insurance market by imposition of a standard contract or by restricting premium rates can exacerbate the problem of adverse selection, and lead to chronic market instability. The only ways to guarantee SB equilibria through regulation are either a) to

impose limits to the possibility of undercutting existing contracts through a minimum insurance requirement; b) directly prohibit the possibility of selecting individual customers.

The conclusions of the theoretical literature are of great practical relevance. Recent reform plans in Germany, Holland and other European countries (see Chapter 7, for the case of the Czech Republic), as well as Latin America, which intended to progressively develop competition among several insurers as a means to indirectly improve service provision, foresaw the creation of a public compensatory fund to manage financial cross-transfers among insurers characterized by different risk pools. In the intention of reformers such cross-transfers should have eliminated the incentive to screen risks, but in several countries their efficacy has so far been limited, given that only imperfect risk adjusters such as age, gender and region of residence could be used as criteria for deciding which contracts to subsidize. The evidence shows that demographic criteria alone are indeed weak predictors of future health expenditure, and they are thus not able to adequately correct the consequences of adverse selection (Van de Ven and Van Vliet, 1992. See discussion of risk and need adjusters in the context of RA mechanisms in Chapter 6).

In countries such as the US that have traditionally relied on private insurers and competition to cover health risks, new forms of regulation are continuously proposed to limit insurers' incentives to risk-select and to respond to the other sources of market failures. One of the more interesting reform proposals was articulated by Diamond (1992), who suggested that insurers (health plans) be required to serve all members of any Health Alliance (whose members would be geographically-defined, or be based on employment). Competition for the market (to serve all the members of any Health Alliance) would be preserved, but competition within each market (to select some members within any Health Alliance according to their risk) would be prohibited.[51]

No country, however, has yet found a completely satisfactory equilibrium between the need to stimulate more competition on the insurers'/purchasers' side, and the need to preserve universal insurance coverage and equity[52].

---

[51] The Clinton's administration initially endorsed Diamond's proposal, but then rejected it after the defeat in the 1992 Senate elections (see also case study in section 6.5.1). So far, in the US the more severe consequences of risk-selection are tamed through Medicare and Medicaid, the government funded insurance programs that cover the elderly, and some of the poor and more vulnerable members of society, who would otherwise not be able to afford to pay a premium commensurate to their health risk.

[52] See discussion of the first trade-off, Chapter 6, section 6.5.1

Annex 2.1 Equity: a Justification for Government Intervention in Health

If we look back in history and read the main documents left by those who first advocated a strong state involvement in the health sector in Europe, we see that the core arguments behind their proposals were in fact rather different from those provided within modern economics. Interestingly, the same applies to education and social security. Thus, it is worth also re-visiting these arguments, as they have profoundly influenced the way public intervention has been shaped.

Equity rather than efficiency concerns seem to lie at the heart of the first proposals for extending insurance coverage, and then guaranteeing universal health insurance. Those committed to universal health coverage believed that by nullifying the "price barrier" it would have been possible to obtain equality of access of all citizens to those services, like education and health, reckoned as fundamental rights to each person[53]. Beveridge, who backed the plan for the establishment of a National Health Service, praised a service that could provide treatment "to every citizen without exception, without remuneration limit and without an economic barrier at any point to delay recourse to it" (Beveridge[54], 1942, p. 162).

Thus, those who argued for a universal health insurance system wanted to eliminate in health the function of market prices in rationing resources according to ability to pay. One could argue that the more immediate way to do that would have been to equalize purchasing power differentials across individuals, by providing targeted subsidies to the poor. By these means, it would also been possible to maintain decentralized, price-based co-ordination mechanisms in the health system. The recent debate on "vouchers" for education, and for some health services, points towards that direction. However, those who created the modern welfare state and National Health Service in the United Kingdom and then in other European Nations opposed these targeted demand-side subsidies as a means to improve access, because they opposed the old British system of assistance based on the Poor Laws, which was precisely based upon means-tested, targeted subsidies. According to its "founding fathers", the new welfare state was on the contrary to be based on universal contributions and benefits (insurance principle), reflecting the association of equals aimed at protection against the principal uncertainties of life. Social stigma, associated with means-tested benefits, had to be eliminated. By using a modern economic terminology, one can say that behind the implementation of

---

[53] Tobin, J. On limiting the domain of inequality. Journal of Law and Economics 1970; 13: 263-277.
[54] Beveridge, W. Social Insurance and Allied Services. London: HMSO, Cmd 6404, 1942.

universal health insurance schemes, there was a life-cycle argument based on equity as well as efficiency: the young and healthy (low risk) accept to contribute to funding social services because they know that in the past as new-borne, and in the future when they will become old and sick, they were and will be able to benefit from subsidized health insurance coverage themselves. At the same time, it is implicitly assumed that such interpersonal and intergenerational cross-transfers can be most effectively implemented at the National level.

A further step, followed by some but not all countries, was to try to guarantee more even levels of care opportunities to everybody, and this necessitated a uniform distribution of physical infrastructures with comparable standards across the country. This last goal was considered exceedingly difficult to achieve within a system based on market or "quasi-market" incentives. Hence, the aim was pursued with the formation of a National Health Service, based not only on universal health insurance, but also on the right to uniform standards of provision across the country. This was first implemented in the UK, then in Scandinavian countries, and then in South Europe (in Italy in 1978).[55]

By contrast with more recent debates and the central focus of modern economic theory, problems of incentives for the armies of civil servants that were created (which included teachers, doctors, nurses, administrators, etc.) were completely overlooked. Hence, having completely eliminated decentralized resource allocation and co-ordination mechanisms, National health services replaced them with central planning. In the UK, after the creation of the NHS in 1948, hospitals and other health providers were nationalized, and the delivery of most health services was organized through public, vertically integrated agencies. Most production decisions, such as volume and mix of inputs, salaries for health workers and other personnel, and investments, were taken at the central level. At the local level, Health Authorities, the public agencies locally responsible for planning and provision of health services, directly managed hospitals. At the same time, total resources for health were directly rationed by the government, who each year determined the global budget for health according to disposable resources and political priorities.

---

[55] Other industrialized countries, among which France, Germany and Canada did not follow the UK path and did not create a National Health Service. They opted for a mixed system, where universal health coverage is guaranteed, but not all health providers are nationalized. These countries substituted risk-related premiums with income-related contributions, but at the same time left a decentralized, although highly regulated, delivery system. By these means, they intended to maintain some private economic incentives on the supply side, for example linking doctors' and providers' payment to the services they actually provide.

Annex 2.2 An introduction to (demand-side) moral hazard

First, consider the subsidization of a certain pharmaceutical product by a private or a public insurance scheme. The subsidy makes the product less expensive at the point of service and may result in a demand-side "moral hazard" effect, depending on the size of the demand and supply elasticity. The point can be illustrated graphically:

Figure 2.16: Demand-side Moral Hazard
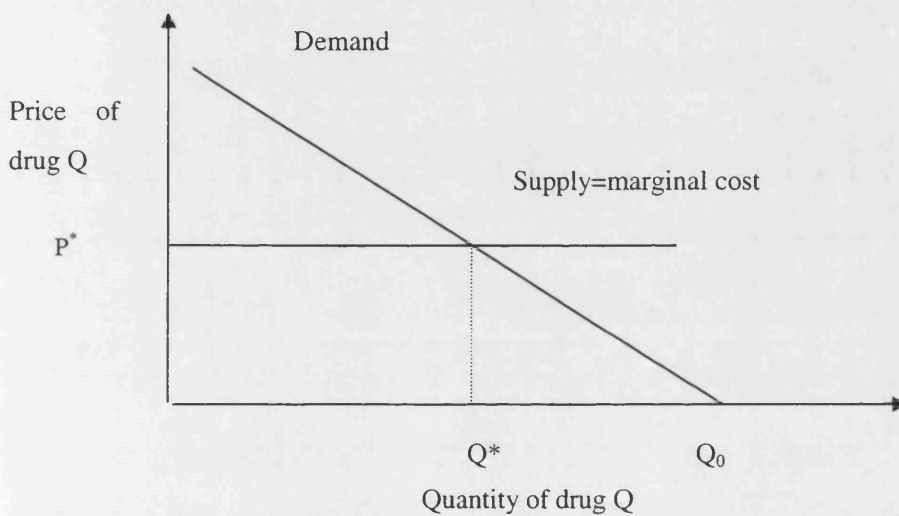
Suppose Q represents the level of consumption of the pharmaceutical product, and that marginal cost (cost of providing each additional unit) is constant, so that supply is infinitely elastic, as in the figure above. Efficiency requires that equilibrium consumption be at the level Q^*, where marginal cost (which, at each level of consumption, is a measure of the value of the additional resources society has to commit to provide an extra-unit of the product) and aggregate demand (the sum of individual demands, which is a measure of social marginal willingness to pay for the product) are equal.

However, if the pharmaceutical product is available for free at the point of service, consumption is increased up to quantity Q_0. In other words, people consume "too much" of the subsidized pharmaceutical product, because it is available for free (in fact some people may still have to sustain transportation costs and other costs not considered here).

In the above situation, a co-payment is able to reduce the level of consumption to a level closer to the optimal level. For example, the following figure shows that by imposing a co-payment equal to M the level of consumption is reduced to Qm, a level closer to the optimal level. Evidence shows (se, for example, the results of the Rand Experiment Group, Manning et al. 1987[56]) that people tend to abuse consumption of certain health services such as pharmaceutical products, outpatient specialist visits or hospital elective care, when they are offered totally for free. In these cases, co-payments can play a role in "screening" demand and make sure that only the more justified claims (in terms of willingness to pay) insist on the scarce resources available for health.

Figure 2.17: The Role of Co-payments in Limiting Over-consumption
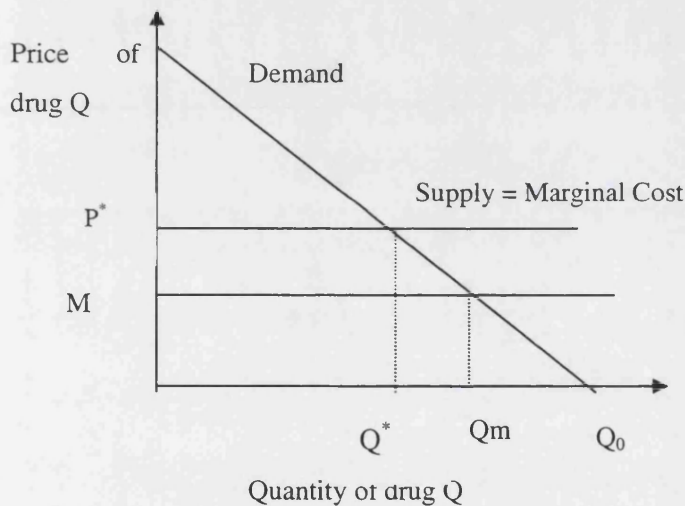
However, quite extensive evidence indicates that for other types of health services demand is not so much influenced by price (empirical studies from developed countries in general estimate quite low price elasticity for medical service –roughly $= - 0.2$ on average). For example, the demand containment effect is likely to be negligible for emergency care. When elasticity is low, the demand containment effect of co-payments is negligible.

---

[56] W. Manning et al. (1987), Health insurance and the Demand for Medical Care: Evidence from a Randomised Experiment, *American Economic Review* 77: 251-277.

Figure 2.18: Inelastic Demand. Service Consumption Level Is not Affected By Co-payments
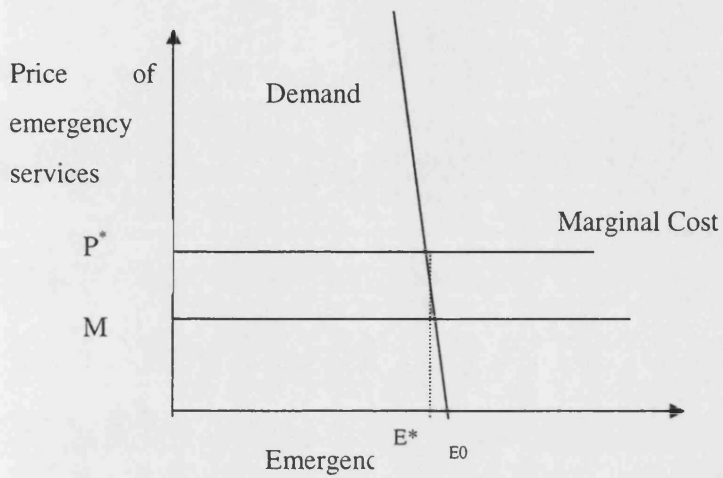
In conclusion, every time co-payments are raised, we are likely to observe some demand reduction, unless demand is completely inelastic. However, the real question is whether co-payments are discouraging unnecessary recourse to treatment, or are they excluding those who are unable to pay?

Consider that the extent of demand for health services and its elasticity are also influenced by information, wealth and cultural attitudes, so that it varies significantly across different socio-economic groups. Thus, individuals who share the same health condition (have the same health needs) and face the same prices, in fact use health services very differently. For instance, poor people in rural areas may not use health services even if a small co-payment is imposed because they are not aware of the benefits they could receive from them, or simply because being in "poor health" is accepted as a normal condition, rather than an accident to be cured. A recent study on Indonesia that investigated into the consequences of the imposition of higher co-payments for health services in a selected group of districts showed that demand elasticity for the poor was much higher than for the rich (Gertler and Molyneaux, 1990).

Chapter 3: Quality and Accessibility of Health Services

3.1 Introduction

This Chapter focuses on the relationship between the dynamics of competition, and quality and accessibility of care. It studies the competitive interaction between service providers in the health sector through a model of vertical differentiation[57], along the lines first developed by Gabszewicz and Thisse (1979, and 1980) and Shaked and Sutton (1982, 1983, 1984). In the first part, the model characterizes the different possible market equilibria when two health providers are competing with each other and are supplying health services of different quality. In the second part, the Chapter discusses the role of public regulation imposing a Minimum Quality Standard (MQS) on the low quality provider, or a universal coverage constraint through or a Price Cap (PC). Our main purpose is to build upon the existing literature on vertically differentiated markets to study the dynamics of competition for the provision of health services, paying particular attention to the impact of competition on quality and accessibility of services. The model presented in this chapter aims to improve upon existing theoretical results (Wauthy, 1996, Tirole, 1988, Ronnen, 1991), by considering the impact of positive quality-enhancing fixed costs on market equilibria, and in comparing equilibria under different forms of competition and regulatory regimes

For all health services that are excludable (approximately 95 percent of the supply of health services) the private sector (not-for-profit and for profit) plays a significant role both in financing and in provision in almost all countries. In developing countries the private sector has a particularly prominent role in provision of primary care and elective secondary care health services (Berman, 1995; Bennett et al., 1997)[58]. For instance, it has been estimated that in India over 80 percent of

---

[57] Vertical product differentiation is defined in the following way: "Given two distinct products, if they were sold at the same price, then all consumers would choose the same one (the highest quality product)" (Shaked and Sutton, 1987). Shaked and Sutton (1983) show that markets where quality-enhancing costs fall primarily on fixed costs are "natural oligopolies", that means they remain concentrated, with few competitors occupying the whole market, even when market size increases (the so-called "finiteness property").

[58] Note that in several countries (particularly sub-Saharan Africa) the not-for-profit (NFP) private sector is as important as the for profit (FP) sector and the government sector in service provision, especially in the poorer areas. There is evidence suggesting that in fact NFP facilities charge higher user-fees than government facilities, and that patients still use them more than government facilities probably because their services are

outpatient contacts take place in the private sector (Yazbeck, 2000). The same was true in all Western European countries until the decade of the 1920s and 1930s, when governments started to play a more significant role in service provision in several places.

Also in the few developing countries where the public delivery system infrastructure is more developed, the evidence suggests that "bypassing" of low-quality public facilities in favour of private facilities is a widespread phenomenon (see Chapter 6), and that service quality has a strong effect on utilization (Alderman and Lavy, 1996).

Understanding market interaction in service provision would provide a rigorous basis to evaluate regulatory policies, such as the imposition of a Minimum Quality Standard, or price regulation, or subsidization and taxation of the private health sector. Such forms of intervention in the health sector are complementary to, but to an extent also alternative, to direct service provision.

Our contribution is complementary to that by Hammer (1998). Hammer highlighted the importance of considering cross price-elasticity between private and public supply and demand in evaluating public interventions. Whilst the latter relationships are characterized by substitutability between the public and private sector, we show that for quality of services the public private interaction is one of strategic complementarity [59].

3.2 The Model

Providers' interaction is modelled as a two-stage game where the players simultaneously choose quality first and then choose prices or quantities (the case of price or Berrand competition in the last stage is analysed in the following, while the case of quantity, or Cournot competition is described in Annex 3.1). Both providers know their competitor's quality level, before the last stage of the game is played[60]. We solve this game by backward induction and characterize the sub-game perfect equilibrium of the game[61].

---

perceived as being of better quality (for Kenya, see Mwabo, Ainsworth, and Nyamete, 1993; for Zambia, see Nakamba, Hanson and McPake, 2002, for Uganda, see Giusti, 2000).

[59] In simple words, qualities are strategic complementaries if an increase in public services' quality makes it convenient for the private sector to increase its quality as well (and vice-versa). We prove that this is actually true in our model.

[60] Our model does not explicitly consider information issues. The hypothesis is that the quality of each

A market where two goods of different quality are supplied is called vertically differentiated if one of the two goods is unanimously preferred to the other if they are sold at the same price (see first footnote at the beginning of this chapter). Considering the perceived qualities of the two goods, low quality, q_l and high quality, q_h, as vectors composed of n characteristics z_{il} and z_{ih} (Lancaster, 1955) where i=1,...,n, the two goods are vertically differentiated if and only if: $z_{il} \geq z_{ih}$ for each i=1,...,n, with strict inequality for at least one i. This property must hold for all potential consumers in the market.

Suppose two health providers are competing to serve a specific market area. For instance, consider the interaction between two clinics in a small town, or between two physicians who are the sole providers in a rural area and compete with each other for patients. For some reason not investigated in the model, one of the two providers is able to initially make a larger investment in quality-enhancing fixed-costs[62]. Once the quality investments are completed, the higher quality provider's services would be preferred by all to the services offered by her competitor, if their services were available at the same price. The higher quality provider has two possible strategies: 1) either she sets

provider's services is known both by the other provider and by patients when the price game is played in the last stage. This may be considered as a first approximation to a more realistic model, which would explicitly consider information problems. Then, quality would be perceived with an error (possibly different for different individuals) and investing in quality may positively influence the provider's probability of being perceived of higher quality.

[61] For a formal definition of sub game-perfect equilibrium, see Fudemberg and Tirole, Ch. 3, p. 92. Sub game perfect equilibrium is a refinement of the Nash equilibrium concept for dynamic games. Sub game equilibria are strategies, which are equilibria not only of the whole game, but also of any proper sub game. The game we consider has just one sub game, which is the price game played in the last stage.

[62] What defines a higher quality health service is an extremely complicated and disputed issue. In considering health quality dimensions, one critical distinction is between those elements of quality that mainly contribute to technical or medical quality and those, such as the quality of ancillary services in hospitals, that are part of perceived quality but do not significantly add to health outcomes. This distinction is not the focus of our analysis.

Another distinction that it is possible to draw is between dimensions of quality that mainly contribute to fixed costs and those that mainly contribute to recurrent costs. The first group of quality variables consists of the quality of equipment, of premises, the level of doctors and nurses' training and, more in general, comprises all the endowment in physical and human capital at each provider's disposal. It is realistic to assume that these dimensions of quality take time to be adjusted, and must be set before the provider decides what prices to charge for services and what volumes of service to deliver. Once investments in physical and human capital are decided, by and large their cost cannot be retrieved (they are sunk in the economists' terminology). The second vector of quality variables, usually considered in the health economics literature under the name of "intensity of treatment" (see Chapter 4), consists of such inputs as time spent with each patient, quality of communication, level of client-orientation, etc. Such dimensions of quality directly depend on nurses' and physicians' effort, can be varied on a daily basis and are part of the variable costs. We assume that such variable quality dimensions are the same for the two providers. In our model, marginal costs are constant, and,

a price low enough to pre-empt the entrance of the other competitor in the market, or 2) she sets a price high enough so that the low quality provider can enjoy a positive market share, by attracting (part of) those who have a lower marginal willingness or ability to pay for quality. In turn, the lower quality provider may cover all the patients that are left out by the high quality provider, or he may set a price and a quality level such that a part of the market remains uncovered (those with the lowest marginal ability/willingness to pay for health services). As we shall see, the range of taste parameters and the shape of the cost function determine the strategies chosen by the two providers and boundaries between the diverse possible equilibrium configurations[63].

3.2.1 Assumptions

Assumption 1 (supply side): When supplying a service of quality q_i, with $q_i \in [0, q^+)$, $i=h$; l and $q_h \succ q_l$, each provider pays for quality-development costs according to the fixed cost function $F(q_i) = q_i^2 / 2$. There is a finite limit to the quality available and each provider supplies just one type of quality.

Assumption 2: Patients' indirect utility from consuming health care services is always greater than zero[64], and equal to:

$V(x,p) = xq_i - p_i$, where p_i is the price of service[65] of quality q_i, with $q_i \in [0, q^+)$, and x represents a taste parameter, uniformly distributed over the interval $(\underline{x}, \overline{x})$ with $\underline{x} = 1$ and with unit density.

Note that the type of health service considered here is indivisible and each patient can have at most one unit. We can think of it as a composite commodity (a "package") comprising several diagnosis-treatment services. The patient can decide not to undergo treatment, but once she does so she has to buy the whole "package".

---

without loss of generality (it would only change all results by a constant), we can set them equal to zero.

[63] There may be a range of parameters corresponding to which more than one strategy is feasible, and in that case to determine the optimal strategy we need to compare the profits levels achieved in correspondence to each of them and choose as an equilibrium that which gives the highest profits (see section 3.2.6).

[64] The characterization of preferences above follows Mussa and Rosen, 1978. Indirect utility is equal to zero if the patient decides not to be treated.

[65] For the services provided by the public provider we can think of p_i as comprising all the co-payments (formal and informal) directly imposed on the patient.

Also note that the hypothesis that $\underline{x}$, the minimum taste parameter, is positive is equivalent to considering on the demand side only those who would search for care, if services were provided for free, and at least at a minimum positive quality level. Imposing the restriction that $\underline{x}=1$ allows us to exclusively focus on one single parameter, $\bar{x}$, as a measure of preferences' diversity, or of total market size. Patients characterized by higher values of x are willing to pay more to access higher quality services, perhaps because they dispose of better information concerning their health status, or because they suffer from a more severe illness. The parameter x can also be interpreted as patients' marginal rate of substitution between income and health expenditure[66]. According to this interpretation, a lower x would correspond to a higher marginal utility of income and to a lower income.

Assumptions 1 and 2 above are sufficient to completely characterize preferences and the technology for the health service in question.

3.2.2 Market Demand

In the private duopoly situation, given a quality and a quantity/or price choice by the two providers, patients can be divided into three groups: those who decide not to purchase any medical treatment, those who purchase the low quality service and those who buy the high quality service. Any of these three demand 'segments' may be empty in equilibrium (for example, if the market is fully covered the first 'segment' is empty). The total size of the potential market is measured by: $\bar{x} - 1$.

Let us denote by x^l and x^m the values of the taste parameters that separate the three groups (which we will denote as 'threshold levels' in the following), i.e., the value of the taste parameters for, respectively, the marginal patient (who derives the same utility from purchasing the low-quality service or from not purchasing), and the indifferent patient (who is indifferent between purchasing the high and the low quality services). From the utility function specified above, the marginal patient's taste parameter is characterized as follows:

---

[66] This interpretation of the parameter x has originally been introduced by Gabszewicz and Thisse (1979,1980), Shaked and Sutton (1982,1983,1984) Bonanno (1986) and Ireland (1987).

$$x^l = \frac{p_l}{q_l} \tag{3.1}$$

Whereas the "indifferent" patient's taste parameter is derived as follows:

$$x^m q_h - p_h = x^m q_l - p_l,$$

$$x^m = \frac{p_h - p_l}{q_h - q_l} \tag{3.2}$$

If $x^l = \frac{p_l}{q_l} \leq 1$, in equilibrium all the market is covered. If $x^l \geq x^m$, in equilibrium all the market

is served by the high-quality provider.

Note that in our model, following Wauthy (1996)[67], the complete or incomplete market coverage is an endogenous result of the price and quality game. More precisely, for any value of the upper taste parameter $\bar{x}$, it is the low quality provider that determines whether or not the market is fully covered in equilibrium. *Ceteris paribus*, the choice of a higher quality/lower price by the low-quality provider will increase the equilibrium degree of market coverage.

3.2.3 Competition in the Last Stage: Bertrand Competition

As it is customary in dynamic games, the game that defines the market equilibrium is solved by backward induction, starting from price competition (Bertrand competition[68]) in the last stage of the game. When the two providers strategically set their prices, they have already determined their qualities, so that qualities can be considered as exogenous parameters in the last stage. We first derive step by step the equilibrium under the hypothesis of incomplete market coverage, in order to clarify the dynamics of the game, and then characterize the range of taste parameters for which such equilibrium holds.

---

[67] Unlike our model, Wauthy (1996) does not consider any quality enhancing costs. In our model, the equilibrium that prevails will depend both on the range of taste parameters, as well as on the cost structure.
[68] Equilibrium quantities and revenues are derived for the case of Cournot competition in Annex 3.1.

When the market is not fully covered and both competitors are active in the market, the demand functions for the high and the low quality services are respectively given by:

$$y_h = \bar{x} - x^m = \frac{\bar{x}(q_h - q_l) - (p_h - p_l)}{q_h - q_l}$$

(3.3), and:

$$y_l = x^m - x^l = \frac{p_h - p_l}{q_h - q_l} - \frac{p_l}{q_l}$$

(3.4)

Providers simultaneously find equilibrium prices by maximizing:

$$\max_{ph} \pi_h^B = (\bar{x} - x^m)p_h$$

(3.5)

$$\max_{pl} \pi_l^B = \left(x^m - \frac{p_l}{q_l} \right)p_l$$

(3.6)

Subject to:

$$p_l \geq 0$$

(3.7)

$$p_l > q_l$$

(3.8)

From the first-order conditions of the above problem, one can derive the following reaction functions:

$$p_h^B = 1/2\left[\bar{x}(q_h - q_l) + p_l^B\right]$$

(3.9)

$$p_l^B = (1/2)\frac{q_l}{q_h}p_h^B$$

(3.10)

The two reaction functions show that, given quality, the two equilibrium prices are strategic complementaries. When the low-quality producer increases its price, the high quality producer is induced to increase its price as well, and vice versa. The two equilibrium prices and market shares are the following:

$$p_h^B = \frac{2\bar{x}q_h(q_h - q_l)}{4q_h - q_l} \qquad (3.11)$$

$$p_l^B = \frac{\bar{x}q_l(q_h - q_l)}{4q_h - q_l} \qquad (3.12)$$

$$y_h^B = \frac{2\bar{x}q_h}{4q_h - q_l} \qquad (3.13)$$

$$y_l^B = \frac{\bar{x}q_h}{4q_h - q_l} \qquad (3.14)$$

In correspondence to the above equilibrium prices and market shares, the 'threshold' taste parameters x^{l^B} and x^{m^B} are equal to:

$$x^{l^B} = \frac{p_l^B}{q_l} = \frac{\bar{x}(q_h - q_l)}{4q_h - q_l} \qquad (3.15)$$

$$x^{m^B} = \frac{p_h^B - p_l^B}{q_h - q_l} = \frac{\bar{x}(2q_h - q_l)}{4q_h - q_l} \qquad (3.16)$$

The constraint that the market be not fully covered in equilibrium requires that: $p_l^B > \underline{x}q_l = q_l$. Substituting the equilibrium price (equation 3.12) in this inequality, one finds the following condition:

$$\frac{p_l^B}{q_l} = \frac{\bar{x}(q_h - q_l)}{4q_h - q_l} > 1 \qquad (3.17), \text{ or:}$$

$$\bar{x} > \frac{4q_h - q_l}{q_h - q_l} \qquad (3.18)$$

Thus, given two quality choices in the first stage of the game, equilibrium is characterized by incomplete market coverage whenever the range of "taste" parameters (or, according to the other

interpretation of the parameter 'x', the range of patients' ability to pay for health services) is broad enough[69]. The precise range of taste parameters where the incomplete market coverage solution holds can be determined only after analysing the first-stage quality game.

3.2.4 Competition in the First Stage: Equilibrium Qualities

Substituting equilibrium prices obtained in the last stage into the revenue function, one obtains:

$$R_h^B = p_h^B y_h^B = (q_h - q_l)\left(\frac{2\bar{x}q_h}{4q_h - q_l}\right)^2 \tag{3.19}$$

$$R_l^B = p_l^B y_l^B = q_h q_l (q_h - q_l)\left(\frac{\bar{x}}{4q_h - q_l}\right)^2 \tag{3.20}$$

Note that for any quality choice: $0 \le q_i \le q^+$ and such that $q_h > q_l$, equilibrium revenues are positive. By differentiating their quality in the first stage of the game, the two providers are able to set prices above marginal cost (which is equal to zero in our case) and earn a positive profit [70]. The two providers simultaneously determine qualities, given the above equilibrium prices p_h^B and p_l^B by maximizing[71]:

$$Max_{q_h}\pi_h^B = R_h^B - F(q_h) = p_h^B(q_h)y_h^B(q_h) - F(q_h) \tag{3.21}$$

$$Max_{q_l}\pi_l^B = R_l^B - F(q_l) = p_l^B(q_l)y_l^B(q_l) - F(q_l)$$

Subject to: $q_i \in [0, q^+)$ and $q_h \ge q_l$.

---

[69] See next section for an intuitive explanation of this result.

[70] Compare this result with the traditional result under Bertrand competition that the only Nash equilibrium of the game is characterized by prices equal to marginal cost (equal to zero in our case).

[71] In the Cournot case (see Annex 3.1), equilibrium qualities are analogously determined, replacing R_h^B, Bertrand equilibrium revenues in the first stage of the game, with R_i^C, Cournot equilibrium revenues.

It is possible to prove that both the high quality and the low quality producers' revenue functions are concave in quality (Motta, 1993), and that the above maximization problem has a unique solution[72], where both providers earn a positive profit. Choi and Shin (1992) show that when there is no fixed or variable cost of quality provision, the equilibrium levels of quality are such that $q_l = (4/7)q_h$.

By adopting a quadratic fixed cost function, $F(q) = q_i^2 / 2$, Motta (1993) shows that a real solution to the maximization problem above is in correspondence to the equilibrium qualities:

$$q_h^B = 0.2533\bar{x}^2 \qquad (3.22)$$

Thus, $q_l^B \cong 0.2q_h^B$, and Choi and Shin's (1992) result that the two equilibrium qualities when the market is not-fully covered are a fixed proportion of each other is confirmed with positive quality-enhancing costs, although q_l^B is now a lower fixed proportion of q_h^B. In correspondence to solution 3.22, constraint 3.18, which must hold in order for the market not to be fully covered in equilibrium, requires that:

$$q_l^B < q_h^B \frac{\bar{x} - 4}{\bar{x} - 1} \qquad (3.23)$$

Recall that $x \in (1, \bar{x})$. Given the above equilibrium qualities, this constraint holds if and only if[73]:

$$0.2 < \frac{\bar{x} - 4}{\bar{x} - 1} \rightarrow \bar{x} > 4.7 \qquad (3.24)$$

---

[72] Being $F_j(q)$ convex, the first order conditions are sufficient conditions for a local maximum. Being $F'''(\ )\geq 0$ also the conditions for a global maximum are satisfied (Ronnen,1991). For the particular quadratic specification of quality-enhancing fixed costs, Motta (1993) also proves that equilibrium levels of quality are stable, as neither competitor has an incentive to leapfrog the other.
[73] In case of Bertrand competition and, as in the case of Choi and Shin (1992), no costs for developing quality, if we denote by $\bar{x}$ the upper limit and by x_{min} the lower limit of the "taste" parameter, Wauthy (1996) showed that the incomplete market coverage prevails in equilibrium whenever $\bar{x}/x_{min} \geq 8,66$.

Hence, we can now give a numerical value to the intuition presented in equation 3.18. If differences in preferences and/or in ability to pay for health services are broad enough, once the higher quality producer has chosen its profit maximizing quality (by positioning itself to serve the upper segments of the market), there is still enough room for the low quality producer to target medium-range sections of the market, and leave out the poorest/or those less willing to pay. Note that the above results hold as long as the providers cannot price-discriminate among patients (first degree price discrimination): they must charge the same price for the same quality service, independently of who is purchasing it.

3.2.5 The Case of Full Market Coverage

The market equilibrium was derived under the assumption of incomplete market coverage. How do the above results change when the low-quality producer decides to cover all the market[74]? Then, the demand faced by the low-quality producer becomes:

$$y_l = x^m - x^l = \frac{p_h - p_l}{q_h - q_l} - \underline{x} = \frac{p_h - p_l}{q_h - q_l} - 1 \qquad (3.25)$$

In case of Bertrand competition in the last stage, we derive the price equilibrium with full market coverage in the last stage of the game by maximizing:

$$\max_{ph}(\bar{x} - x^m)p_h \qquad (3.26)$$

$$\max_{pl}(x^m - 1)p_l \qquad (3.27)$$

$$s.t. p_l \geq 0 \qquad (3.28)$$

$$p_l \leq q_l \qquad (3.29)$$

---

[74] Tirole (1988, p. 96) was the first one to characterize equilibria with full market coverage. He made the hypothesis of constant variable costs and no fixed costs. He also failed to recognize (Wauthy, 1996) the existence of a range of parameters where a corner solution prevails.

The interior solutions of such maximization are as follows:

$$p_h^* = \frac{(q_h - q_l)(2\bar{x} - 1)}{3} \qquad (3.30)$$

$$p_l^* = \frac{(q_h - q_l)(\bar{x} - 2)}{3} \qquad (3.31)$$

In correspondence to the above equilibrium prices, the constraint that p_l is non-negative requires that $\bar{x} \geq 2$. If $\bar{x} \leq 2$, the high-quality provider pre-empts the market. In other words, the size of the market is too "narrow" (or, preferences are not differentiated enough) to allow more than one provider to profitably enter the market [75].

The constraint that the market be fully covered $(p_l \leq q_l)$ requires that:

$$\bar{x} \leq \frac{2q_h + q_l}{q_h - q_l} \qquad (3.32)$$

Recall that the incomplete market coverage required that $\bar{x} > \dfrac{4q_h - q_l}{q_h - q_l}$.

Hence, there is an intermediate range of the upper-limit taste parameter $\left(\dfrac{2q_h + q_l}{q_h - q_l} < \bar{x} \leq \dfrac{4q_h - q_l}{q_h - q_l} \right)$, equivalent to: $q_h^0 \dfrac{\bar{x} - 4}{\bar{x} - 1} < q_l^0 \leq q_h^0 \dfrac{\bar{x} - 2}{\bar{x} + 1}$ (3.33),

where prices obtained under the hypothesis of full-market coverage $\left(p_l^*, p_h^* \right)$ lead to an equilibrium characterized by incomplete market coverage (they are therefore inconsistent), whilst optimal prices p_l^B, p_h^B, obtained under the hypothesis of incomplete market coverage, in fact lead to fully cover the

---

[75] In this case equilibrium prices would be:

$$p_l^M = 0$$

$$p_h^M = q_h$$

market (constraint 3.8 is binding)[76]. Within such interval, a "corner solution" for p_l prevails (which we will denote by p_l^0). p_l^0 is set exactly equal to q_l, the level at which the marginal patient is just indifferent between purchasing the low-quality service and not purchasing, and all the rest of the market is covered. Let us look at the equilibrium prices and market shares in correspondence to such corner solution. Straightforward computations show that:

$$p_l^0 = \underline{x}q_l = q_l \tag{3.34}$$

$$p_h^0 = \frac{1}{2}\bar{x}(q_h - q_l) + \frac{1}{2}q_l \tag{3.35}$$

Equilibrium market shares are as follows:

$$y_h^0 = \frac{1}{2}\bar{x} + \frac{1}{2}\frac{q_l}{q_h - q_l} \tag{3.36}$$

$$y_l^0 = \frac{1}{2}\bar{x} - \frac{1}{2}\frac{2q_h - q_l}{q_h - q_l} \tag{3.37}$$

3.2.6 Equilibrium Qualities When the Market is Fully Covered

Let us now turn to the quality game in the first stage. As it was the case under incomplete market coverage (equation 3.21), optimal qualities are found by substituting equilibrium prices and market shares under the assumption of complete market coverage in the revenue function of both competitors, and by maximizing:

$$Max_{q_h}\pi_h = p_h(q_h)y_h(q_h) - q_h^2/2 \tag{3.38}$$

$$Max_{q_l}\pi_l = q_l y_l(q_l) - q_l^2/2$$

---

[76] The existence of a corner solution in between the unconstrained prices with uncovered and covered market was first recognized by Wauthy (1996). Note that moving from an incomplete to a full-market coverage situation changes the price elasticity of the demand faced by the low quality producer. Until the market is fully covered, by reducing its price, the low quality producer is able: 1) to gain market shares vis-à-vis the high quality producer; 2) to increase the size of the market, which is covered in equilibrium. Once the market is fully covered, competition is limited to market shares. Under full-market coverage demand is more rigid, and

Subject to: $q_i \in [0, q^+)$, $q_h \geq q_l$

The optimal qualities will be different in case the interior (**case a.** below) or the corner solution (**case b**) prevails in the last stage of the game.

a. When prices are set according to equations (3.30 and 3.31) in the last stage of the game, x^m, the value of the taste parameter separating those who purchase the high-quality services from those who purchase the low-quality service, is independent of q_l and q_h, and equal to: $x^m = \dfrac{1 + \bar{x}}{3}$. The low quality provider's profit function is decreasing in quality[77], and therefore the unconstrained equilibrium quality q_l would be set equal to $q_l^* = q_{min} = 0$ (in parallel, q_h would be equal to:

$q_h^* = \dfrac{(2\bar{x} - 1)^2}{9} > 0$, independently of q_l). However, in correspondence to $q_l^* = 0$, constraint 3.29 is never satisfied for $p_l \geq 0$, regardless of the value taken by the taste parameter. Hence, the equilibrium qualities with unconstrained choice of prices under full market coverage, can never be sustained as an equilibrium under our characterization of the taste and technology (cost) parameters.

b. In correspondence to the corner solution for prices (equations 3.34 and 3.35 above), the computation of the optimal qualities is analytically complicated. However, it is possible to prove (see Appendix 3.2) that the system of First Order Conditions [78] is characterized by a unique solution for $\dfrac{q_l}{q_h} \leq 1$, defined for any $2 \leq \bar{x} < 4.7$. Such solution finds a lower bound and can be approximated by the following expression:

---

the equilibrium price is set at a comparatively higher level.

[77] That is because market shares are not influenced by quality, whereas p_l is inversely related to q_l:

$$\partial p_l^* / \partial q_l = \partial\left[\dfrac{(q_h - q_l)(\bar{x} - 2)}{3}\right] / \partial q_l < 0.$$

[78] The first order conditions lead to the following system of two equations (see Appendix 3.2):

$$\left(\bar{x} - 1 - 2q_l^0\right)\left(q_h^0 - q_l^0\right)^2 - q_h^{0^2} = 0$$

$$\left(\bar{x}^2 - 4q_h^0\right)\left(q_h^0 - q_l^0\right)^2 - q_l^{0^2} = 0$$

The above system has real solutions, provided $\bar{x} > 1 + 2q_l^0$ and $\bar{x} > 2\sqrt{q_h^0}$.

84

$$q_l^0 = \frac{2(\bar{x}-2)}{(\bar{x}^2+4\bar{x}-4)}q_h^0$$

(3.39)

Note that solution 3.39 satisfies constraint 3.33, necessary for the corner solution to prevail in the last stage of the game, whenever $2 < \bar{x} < 4.7$. [79] By contrast, outside the range of the upper taste parameter $\left(\frac{2q_h + q_l}{q_h - q_l} < \bar{x} \leq \frac{4q_h - q_l}{q_h - q_l}\right)$, choosing the corner solution $p_l^0 = q_l$ cannot be part of a sub-game perfect equilibrium[80].

Also note that the above equilibrium quality ratio $\dfrac{q_l^0}{q_h^0}$ is increasing in $\bar{x}$, the upper limit in the range of taste parameters, for any $\bar{x} < 4.7$. When $\bar{x} = 2$, the threshold value below which only one provider is left in the market, $q_l^0 = 0$. As $\bar{x}$ tends to the value where the incomplete market solution would prevail (equal to 4.7 in our example), $\dfrac{q_l^0}{q_h^0}$ tends to the same ratio as the optimal qualities in case of incomplete market coverage (value equal to approx. 0.2 in our example).

Let us summarize the characterisation of equilibria under the hypothesis of full market coverage. By solving the game by backward induction, we have first proved that an unconstrained choice of equilibrium prices and qualities is not able to sustain an equilibrium with complete market coverage. Only when prices are set according to the "corner solution" in the last stage of the game ($p_l^0 = q_l$), a

---

[79] In fact, the constraint $q_h^0 \dfrac{\bar{x}-4}{\bar{x}-1} < q_l^0$ with $q_l^0 = \dfrac{2(\bar{x}-2)}{(\bar{x}^2+4\bar{x}-4)}q_h^0$ is satisfied only for $\bar{x} < 4.5$ and not for $\bar{x} < 4.7$. This is because q_l^0 is a lower-bound approximation of the exact solution. If we used the latter, we would find that when $\bar{x} = 4.7$ (the threshold value of the taste parameter in correspondence to which the market becomes fully covered), the equilibrium relative qualities are $q_l \cong 0.19 q_h$, identical to those found in case of uncovered-market price equilibrium in the last stage. $\left(0.19 = \dfrac{q_l^0}{q_h^0} = \dfrac{q_l^B}{q_h^B}\right)$.

[80] In order to be a sub-game perfect equilibrium, a strategy must be a Nash equilibrium in any sub-game. The last stage of the game is a sub-game. Outside this intermediate range of parameters the strategy of setting $p_l^0 = \underline{x}q_l = q_l$ would be an "empty threat" on the part of the low quality provider in the lat stage of the game, even if it could be a Nash equilibrium for the whole game.

quality level q_l^0 and q_h^0.\ in correspondence to which the full market equilibrium can be sustained, if the upper taste parameter $\bar{x}$ is equal to: $2 < \bar{x} < 4.7$.

In summary, the set of sub-game-perfect equilibria is dependent on the value assumed by the upper taste parameter, and is characterized as follows:

Proposition 1:

When $\bar{x} > 4.7$ the equilibrium with incomplete market coverage prevails. The lower quality provider sets a quality level, which is a set proportion of the higher quality level provider.

When $2 < \bar{x} \le 4.7$ the equilibrium characterized by full market coverage prevails. The low quality provider sets its price in correspondence to the corner solution $p_l = q_l$. Within this range of the taste parameter, quality q_l increases as preferences become more diversified ($\bar{x}$ increases). The quality ratio approaches the incomplete market coverage equilibrium as $\bar{x} \to 4.7$.

When $1 < \bar{x} \le 2$ the market is pre-empted by the higher-quality provider. In this case, p_h is set exactly equal to q_h.

3.2.7 Introduction to the Analysis of the Impact of the Imposition of a Minimum Quality Standard (MQS)

In this section we analyse what are the implications of the above analysis for the assessment of a minimum quality standard (MQS) policy that imposes a floor on q_l, the quality of the low-quality provider. The impact will depend on whether a incomplete or a complete market coverage equilibrium prevails before the MQS is imposed.

i) **Incomplete market coverage.** As indicated by Ronnen (1991), in case of Bertrand competition in the last stage, starting from a situation of incomplete market coverage the imposition of a MQS on the low-quality provider always bears a positive welfare enhancing effect. A MQS policy that increases the equilibrium quality chosen by the low quality producer (q_l^*) leads to an increase in:

- market coverage;

- the share of the market taken by the high quality producer;
- the level of quality chosen by the high quality producer.

Then, consumer surplus is unequivocally positively affected, and so does welfare (Ronnen, 1991, pp. 497-498, see also common results under Bertrand and Cournot competition in the Annex). The above result highlights an extremely important, and yet neglected role of a MQS regulation: the latter does not only serve the purpose of enhancing service quality for those that already utilize health services, but may also be a tool for safeguarding, or enlarging market participation.

ii) Full market coverage. Suppose that initially $q_l^0 = \dfrac{2(\bar{x}-2)}{(\bar{x}^2 + 4\bar{x} - 4)} q_h^0$ and that a MQS is imposed

on q_l^0. Up to the quality threshold in correspondence to which constraint 3.33 is binding, the low quality provider will continue to choose the corner solution for prices (namely $p_l^* = p_l^0 = q_l$) and progressively raise quality standards to comply with the regulation. However, beyond a certain threshold (where $q_l = q_h \dfrac{\bar{x} - 2x_{min}}{\bar{x} + x_{min}} = q_h \dfrac{\bar{x} - 2}{\bar{x} + 1}$), the low quality provider will switch to the unconstrained choice of prices (according to equation 3.31). Thus, the imposition of a MQS will maintain the equilibrium characterised by full market coverage, improve the level of quality produced by the low quality provider and by these means increase consumers' welfare, but it would also lead to lower profits for the low-quality provider and to higher prices charged by both providers (see Appendix 3.2 for a formal proof of this last result), and by these means reduce welfare. The net effect will depend on the relative evaluation of price-quality variations by consumers.

Also note that if the government cannot subsidize the low-quality provider and if the Minimum Quality Standard is set too high, profits for the low quality producer may become negative and then only one provider would be left in the market. In this latter case, the MQS would lead to a greater concentration in the market and higher prices (lower coverage).

The analysis in this section provides only the first step towards drawing the full welfare implications of the impact of different forms of government intervention, including subsidization of the private sector, price and quality regulation, or direct provision. A full welfare analysis should be based on the specification of a public welfare function, with quality and accessibility of services among its arguments. We leave such further developments to subsequent research.

3.3 Conclusion

In this chapter we analysed the market equilibrium in a vertically differentiated market where health providers compete on price and quality of care. We indicated that market equilibria can be characterised by full or partial coverage depending on the characterisation of patients' preferences, and of quality-enhancing costs. The above analysis suggests that when preferences for health services are relatively uniform, equilibria characterized by full market coverage would prevail, whereas, as the range of taste parameters broadens, equilibria characterized by incomplete market coverage progressively emerge. By comparison with previous models characterised by linear quality-enhancing costs, our analysis also suggests that incomplete market coverage equilibria are more likely when the structure of these costs is characterized by decreasing returns to scale.

The above model also indicated that the anticipation of the intensity of competition determines how 'distant' from each other providers set their equilibrium quality levels. Under full market coverage, competition in prices is likely to be harsher, because there is no more "free space" to occupy and the two competitors compete against each other for their relative market shares. Ex-ante, this has an impact on the determination of optimal qualities, because, anticipating harsh price competition in the second stage (which in the long-run could threaten to drive both competitors' profits down to zero), in the first stage providers move their quality further apart from each other. By so doing they make their services less substitutable of each other, and thus they relax price competition in the last stage. In this case the imposition of a MQS leads to higher quality, but it can also produce higher prices and in the limit, crowd out low quality providers from the market.

As the range of taste parameters broadens, the quality set by both providers increases $\left(\partial q_i / \partial (\bar{x} - x_{\min}) \geq 0\right)$. The quality chosen by the low-quality provider increases more than proportionally until a certain threshold is reached (which is inversely related to the "steepness" of the cost function), after which it remains as a fixed proportion of his competitor's quality $\left(q_l = 0.19 q_h\right)$. The intuitive explanation is that, once the size of the market is large enough or preferences for health services are diversified enough, competitors do not need to differentiate their qualities as much in order to relax price competition, because there is room to always cover untapped segments of the market, without taking those away from the competitor. Thus, in correspondence to equilibria characterized by incomplete market coverage, quality levels tend to be closer to each other. In case

of incomplete market coverage in equilibrium, the imposition of a MQS regulation unequivocally leads to positive welfare effects.

Thus, building upon Wauthy's (1996) analysis (Wauthy (1996) adopted a different specification of the quality-enhancing cost function), our model confirmed his principal result, although the range of taste parameters where the different solutions prevail is obviously different. More specifically, comparing our model to Wauthy (1996), one can see that explicitly considering positive (fixed) costs:

- Broadens the range of taste/income parameters in correspondence to which market equilibria characterized by incomplete coverage prevail.

- Narrows the range of taste/income parameters in correspondence to which equilibria characterized by complete market coverage prevail. At the same time, by contrast with Wauthy's results, we showed that the full-market coverage equilibria can only be maintained under our specification of the costs parameters only if the low-quality provider sets his price at a level that makes the marginal patient indifferent between purchasing or not.

- Leaves unchanged the set of taste parameters in correspondence to which the market is pre-empted by the high quality provider.

What are the predictions of the above model? There are several of them, summarized here. Focusing first on the relationship between quality and accessibility of services, we can underscore the following points:

- Interpreting the health systems' historical evolution in light of the above analysis, one can observe that in the first phases of the epidemiological transition and health systems' development, preferences for health care tend to be more similar across the population, as incomes and expectations are also lower and less dispersed. In such a situation our analysis predicts that all patients will be served by the private providers, but quality differential will be more pronounced (perhaps with untrained providers serving poorer patients). By contrast, as people's expectations increase and become more diversified, quality of care will improve, but at the same time private providers will increasingly leave out the lower segments of the population. Our analysis predicts that the market equilibrium will be characterised by an upper tier of the population given high-quality services, a middle tier served given lower quality services, and a lower tier excluded from the market.

- Considering different health services at any given point in time, another question of interest is the following: "For which services different equilibria are more likely to prevail?" As we indicated, the key parameters are relative to preferences and technology. For more costly and sophisticated services (such as tertiary care services) quality-enhancing costs are likely to increase more rapidly, and the distribution of preferences to be more dispersed across the population. By contrast, the hypothesis of similar tastes across the population is more realistic for basic or routine care. Then, our model predicts that for the former group of services a more significant share of the population will be left out, because of the equilibrium choices of prices and quality. However, it also predicts that the problem of controlling quality is more urgent for basic rather than more complicated services, because providers will tend to diversify more quality levels for basic services, to decrease the intensity of competition.

To summarize the key characteristics of the oligopolistic equilibrium, we can recall the following results:

- For any quality choice: $0 \leq q_i \leq \infty$ and such that $q_h > q_l$, equilibrium profits are positive. By differentiating qualities, the two providers are able to set prices above marginal cost (which is equal to zero in our case) and earn a positive profit.
- Given quality of service provision, as the low-quality provider increases its prices, the high-quality provider increases its prices as well, and vice versa. Thus, the two equilibrium prices are strategic complementaries (by contrast, quantities are strategic substitutes; see Hammer, 1998).

Finally, we are also interested in studying the reaction of the higher quality producer to exogenous changes in the quality set by the low quality producer, starting from the market equilibrium level, and under the hypothesis of incomplete market coverage:

- Increasing the quality of the low quality provider, q_l, in the short-term (when q_h is fixed) leads to:

    a. decreases in the prices charged by the high-quality provider, p_h;

    b. decreases profits for the high-quality provider. Note that the effects are not linear. The decrease in p_h caused by an increase in q_l is stronger the lower is the initial quality offered by the low-quality provider (the second derivative of p_h with respect to q_l is positive, under both Bertrand and Cournot competitive settings).

c. The two qualities of service are strategic complementaries, which means that the increase in the quality of one of the two providers leads to an increase in the other provider's quality.

Annex 3.1 Corner-Solution: Equilibrium Quality Levels

When the corner solution prevails for prices ($p^0{}_l = q_l$), from maximization problem 3.38 one can derive the following First Order necessary Conditions:

$$(\bar{x} - 1 - 2q_l)(q_h - q_l)^2 - q_h^2 = 0 \tag{3.56}$$

$$(\bar{x}^2 - 4q_h)(q_h - q_l)^2 - q_l^2 = 0 \tag{3.57}$$

with $\bar{x} > 2$, $q_h \geq q_l \geq 0$

To solve the above system from the first equation 3.56, derive:

$$(q_h - q_l)^2 = \frac{q_h^2}{(\bar{x} - 1 - 2q_l)} \tag{3.58}$$

Substitute $(q_h - q_l)^2$ from equation 3.58 in equation 3.57:

$$4q_h^3 - \bar{x}q_h^2 = 2q_l^3 - (\bar{x} - 1)q_l^2 \tag{3.59}$$

Define $f = \dfrac{q_l}{q_h} \leq 1$. Divide each member of 3.59 by q_h^3. Then one obtains:

$$q_h = \frac{(\bar{x} - 1)f^2 - \bar{x}^2}{2f^3 - 4} \tag{3.60}$$

Now, divide each member of equation 3.57 by q_h^3:

$$q_h = \frac{\bar{x}^2 + \bar{x}^2 f^2 - 2\bar{x}^2 f - f^2}{4(1 + f^2 - 2f)} \tag{3.61}$$

The last two equations 3.60 and 3.61 can be combined, thus obtaining a cubic equation in f :

$$z(f) = f^3(1 - \bar{x}^2) + 2f^2(\bar{x} + \bar{x}^2 - 1) + f(4 - \bar{x}^2 - 4\bar{x}) + 2(\bar{x} - 2) = 0 \qquad (3.62)$$

with $f = \dfrac{q_l}{q_h} \leq 1$

Let us study the function z(f):

1.

    a.   when f=0, z(f) =$2\bar{x}$ -4 > 0, because $\bar{x}$ >2;

    b.   when f=1, z(f) =- 1

    c. $\lim_{f \to \infty} z(f) \to -\infty, because 1 - \bar{x}^2 < 0$. However, we are not interested in values of

$f = \dfrac{q_l}{q_h} > 1$.

    Thus, for 0<f<1, there is at least a solution z(f) = 0, for each $\bar{x}$.

2. Let us compute the first and second derivative of z(f):

$$z'(f) = 3f^2(1 - \bar{x}^2) + 4f(\bar{x} + \bar{x}^2 - 1) + (4 - \bar{x}^2 - 4\bar{x}) \qquad (3.63)$$

$$z''(f) = 6f(1 - \bar{x}^2) + 4(\bar{x} + \bar{x}^2 - 1) \qquad (3.64)$$

We want to find the relative max and min of the cubic equation in f: z'(f)=0.

$$f = \frac{-2(\bar{x} + \bar{x}^2 - 1) \pm \sqrt{4(\bar{x} + \bar{x}^2 - 1)^2 - 3(1 - \bar{x}^2)(4 - 4\bar{x} - \bar{x}^2)}}{3(1 - \bar{x}^2)} \qquad (3.65)$$

For $\bar{x}$ >2: $(\bar{x} + \bar{x}^2 - 1)$>0, (1-$\bar{x}^2$)<0, $(4 - 4\bar{x} - \bar{x}^2)$<0.

The product under square root ($3(1 - \bar{x}^2)(4 - 4\bar{x} - \bar{x}^2)$) is positive, and therefore the whole expression under square root is smaller than the square of the term outside the root, which implies that the numerator of expression 3.65 is negative for each $\bar{x}$. The denominator is also negative, and

so the two solutions of 3.65 are both positive. In turn, this entails that the cubic expression 3.62 has its relative max and min in correspondence to f>0. Let us compute the derivatives in:

f=0: $z'(f) = \left(4 - 4\bar{x} - \bar{x}^2\right) < 0$

f=1: $z'(f) = 3$

Thus, between $f \in (0,1)$, the function $z(f)$ has a relative min, which implies that the solution to equation 3.62 is unique (there is only one point where the graph of $z(f)$ crosses the horizontal axis between 0 and 1. Let us see the graph of $z(f)$:

Figure 3.1: Equilibrium Quality Ratios in Correspondence to the Corner Solution for Prices
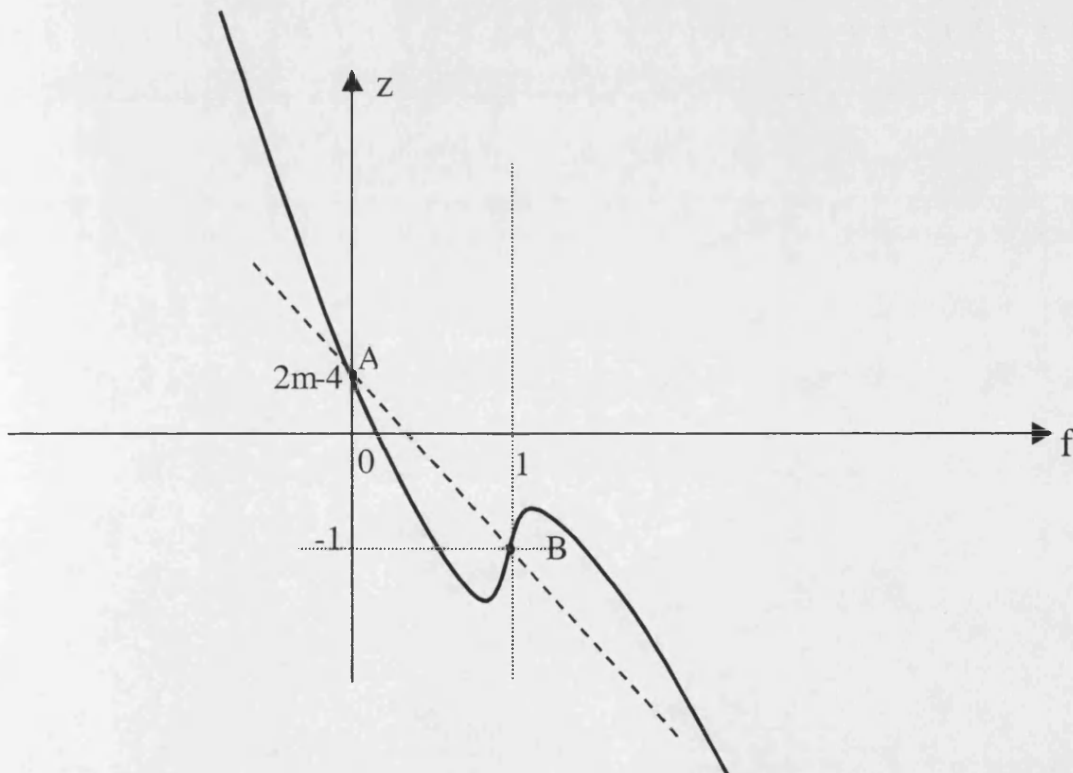
Legenda: $m = \bar{x}$ in the text.

The unique solution f^*: $z(f)=0$ is comprised between

$$\frac{2(\bar{x} - 2))}{\left(\bar{x}^2 + 4\bar{x} - 4\right)} < f^* < \frac{2(\bar{x} - 2))}{(2\bar{x} - 3)}$$

The above two values are the intercepts in f=0 of the tangent to the cubic

($z^T = 2(\bar{x} - 2) + f(4 - \bar{x}^2 - 4\bar{x})$) and of the segment AB (dashed blue segment in the above

picture).

Annex 3.2 Public-Private Interaction in Health

In this Annex, some of the possible channels of interaction between public and private provision are investigated within a simple graphical supply-demand framework.

Whenever the public sector offers or subsidizes health services that the private sector already provides, this is most likely to influence the way the latter operates. There may be "crowding-out" or "crowding in" effects, as well as variations in quality of services offered by the private sector.

Recognizing the role of the private sector would change the perspective according to which priorities are defined in the public sector. As Hammer (1997, p. 48) writes: "Government investment, like any other government intervention, should be justified in terms of the social benefit the project will have over and above that which would occur without public sector involvement. For any investment opportunity, the focus of analysis should be on the difference between social and private benefits-not on the costs and expected returns to private goods themselves". [81]

Focusing on service provision, note that new evidence is beginning to emerge from studies in developing countries showing the significance of public-private interaction. For example, in a multi-country study in several African sub-Saharan countries, Alderman and Lavy (1996) found a significant reduction in the use of private facilities determined by greater availability (measured as the reduction of distance from the closest facility) or higher quality of public facilities (mainly measured as improved drug availability). In a study on Indonesia Gertler and Molineaux (1995) showed that public and private fees are correlated and that demand changes depend on both prices. Alderman and Gertler (1989) used sensitivity analysis to study the effects on demand for Pakistan public and private facilities of changing public service fees. Again, the estimated cross-price elasticity between public and private services are significant.

---

[81] As another example, by contributing to subsidize health plans for high risk people or to directly finance medical treatments for most catastrophic illnesses, the government may bring about a greater welfare gain than by providing primary care services that are already accessible in the private sector totally for free (see analysis of adverse selection in Chapter 2, and Chapter 6). Note that catastrophic care services in fact tend to occupy low levels in priority scales based on Medical Intervention Cost-Effectiveness analysis (see Annex 6.4, and Hammer and Berman, 1995).

The change in the equilibrium will be influenced by demand and supply elasticity[82], and by demand cross-price elasticity, which measures how demand for a certain good or service is affected by the change in the price of another good or service. If the two goods are complementary, cross-price elasticity is negative, if they are substitutes, it is positive[83].

In order to illustrate the importance of measuring also cross-price elasticity across different medical services, suppose that in country X the government decides to impose co-payments on pharmaceuticals prescribed during outpatient care visits, while those prescribed in hospitals remain for free. Then, we may suspect that people would search admission in hospitals just to receive free prescriptions.

To measure such substitution effect between drugs received in outpatient and inpatient care, the relevant concept is that of demand cross-price elasticity. If Q_o is volume of pharmaceuticals demanded outside the hospital, Q_i, is volume of pharmaceuticals demanded inside the hospital, P_0 is the price of drugs at the point of service outside the hospital and P_i is the price at the point of service inside the hospital (equal to zero in our example), and m individual income, we have two demand functions for pharmaceuticals for each consumer:

$Q_o = f (P_0, P_i, m)$

$Q_i = f (P_0, P_i, m)$

The percentage change in demand of drugs inside the hospital, due to a 1% increase in prices of pharmaceuticals purchased outside the hospital is measured by the cross-price elasticity of Q_i with respect to P_0

$$e_{Q_i P_0} = \left| \frac{\partial Q_i}{\partial P_0} \frac{P_0}{Q_i} \right|$$

The greater is the above elasticity, the greater will be the substitubility between pharmaceuticals purchased inside and outside the hospital. In case of perfect substitubility, any increase in the price

---

[82] Demand (price) elasticity for any particular good or service is the percentage change in market demand for that good or service determined by a 1% percent change in its market price. Supply elasticity is similarly measured.

[83] The absolute value of such elasticity conveys information on how "close" to each other the markets for the two goods or services are, and how strong the influence of changes in one market on the other market equilibrium is likely to be.

of drugs purchased outside the hospitals above the price paid in the hospital, will result in a complete collapse of the "market" for drugs outside the hospital and no revenue would be collected through the higher co-payments.

Considering the supply elasticity is equally important. Let us take an extreme example, and suppose that government chooses to provide a demand subsidy for a health service which is already in short supply (there is no unused capacity in place). Suppose also that at least in the short term doctors and equipment cannot be reallocated from other services to increase the supply of that service.

Figure 3.2: Inelastic Supply- Subsidy Leads to Price Increase
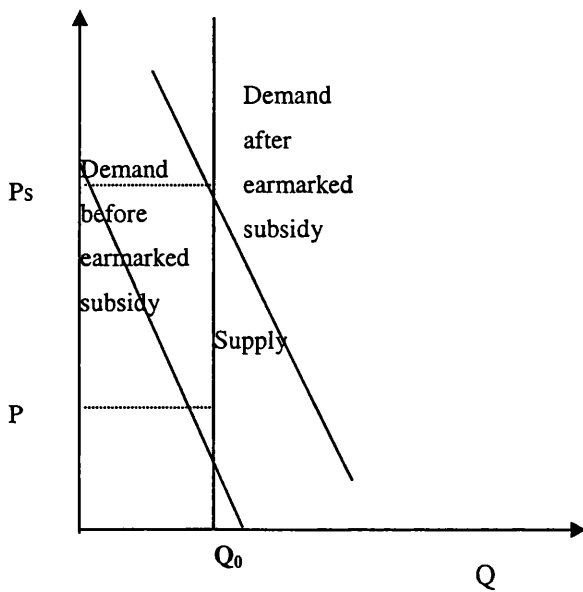
In the above situation the increased demand causes only a price increase, equal to the amount of the subsidy. There is no adjustment in the quantity supplied. Moreover, if the subsidy is given only to a sub-group of the population (for example, trough a social insurance scheme exclusively for formal sector employees), the price increase may make the subsidized services unaffordable to the rest of the population.

Hence, the best public intervention on the supply side is different in case when private supply is elastic and when it is inelastic. When it is inelastic, as in the above example (which is generally true for more sophisticated health inputs and more complex services), the only way of changing the

market equilibrium supply is to directly supplement it with public delivery. There is little crowding out of private suppliers, even with significant changes in equilibrium prices. A totally different course of action ought to be followed when supply is elastic, as it might be the case with primary care services. If supply is elastic, public delivery leads to strong crowding out of private suppliers, such that in extreme situations (as illustrated in the figure below) total supply may not change despite public intervention. In these situations, it is also important to measure the degree of substitutability between private and public services, and the way to measure is to estimate the demand cross-price elasticity between public and private services. If such elasticity is not significant, it means that the market for private and public care may be considered as totally independent markets. Even if private supply is elastic, there would then be no crowding out because of increase of government supply.

Figure 3.3: The Impact of a Health Information Campaign, and the Crowding-out of Private Providers
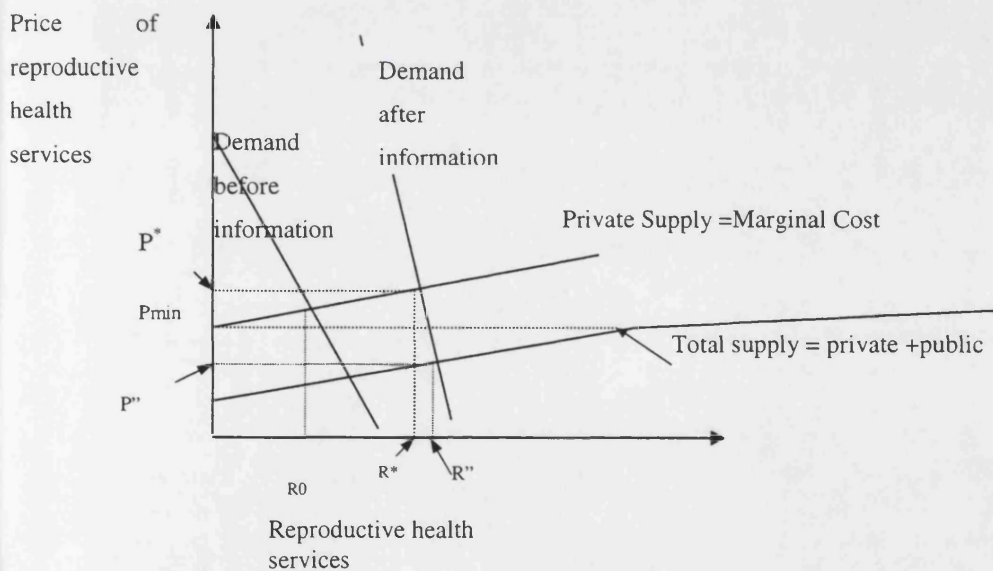
In the figure above, the information campaign shifts demand for reproductive health services significantly outwards. The new equilibrium is characterized by price P* and equilibrium quantity R*. Considering that price too high, the government might want to provide extra reproductive services at a lower price in the public sector. However, in the above figure (1) there is perfect substitubility between public and private services and (2) private supply is very elastic (sensitive to

price). Due to (1) the private sector has to lower its prices to maintain market shares. If supply is elastic, as in the above example, public supply crowds out private supply (the private sector does not provide any service for a price lower than Pmin). In the new equilibrium (P"" and R''), the volume of reproductive health services is not significantly different from the private equilibrium, and cheaper public services simply substitute for more expensive private ones.

Chapter 4: Health Reforms

4.1 Internal market reforms

In this and the following chapters we will focus on the resource allocation, purchasing and payment system (briefly RAP) components of the internal market reforms that, following the UK and the US examples, took place in several lower income countries over the decade of the '90s. In this Chapter we will provide a broad overview of the reforms, their rational and main components. We will also introduce the conceptual framework, based on the notion of trade-offs, which informs our reading of the impact of the reforms in this and the next chapters. Chapter 5 is dedicated to a more in depth analysis of payment systems (PS) and contracts between purchasers and providers in health. Chapter 6 looks more in detail and presents the empirical evidence on the impact of the reforms from an equity perspective. Chapter 7 provides an empirical analysis of the impact of the reforms in the Former Socialist Economies, and Chapter 8 presents a study of the impact of the reforms in Georgia.

As we anticipated in Chapter 1, the internal market or managed competition reforms, which took place in several industrialised and middle income countries from the beginning of the decade of the '90s, emphasized the role of competition among providers as a means to achieve more efficient and client-oriented delivery of publicly financed health services, and that of the health purchaser as an agency be interposed between patients and providers to "discipline" the market. Purchasers would "organize specific types of health care for a designated population (whether defined by geography, employment type or voluntary enrolment)" (Rice and Smith, 2000, p. 1). The motivation of the purchasing role was justified as follows: by creating purchasers charged with the role of screening demand and of bargaining with providers on patients' behalf it would be possible to enhance demand's sensitivity to quality and/or cost, and still not renounce universal insurance coverage where such coverage was already in place. These institutional purchasers would be better informed, and have a greater bargaining power than patients did. Moreover, their purchasing power, unlike patients', could be easily equalized (by financing/subsidizing them according to a capitation formula, for example).

According to the theory behind the reforms (Enthoven, 1985), the government should have retained only the role of (partially) financing specific health services, and of externally regulating quality of care.

In the UK, the reforms became known as "internal or quasi market" health reforms because they intended to replicate "market-type" mechanisms for the purchasing and payment of publicly funded health services. [84] In the UK, the reforms had three main components:

a) split of Health Authorities (HAs), previously the local authorities in charge of hospitals and all other health service providers, into provider and purchaser units;

b) corporatisation of hospitals, progressively turned into Hospital Trusts with separate budgets and semi-independent management, mainly funded according to contracts set with HAs and GP Fund holders;

c) creation of GP Fund holders (GPFHs), alongside Health Authorities, on the purchaser side[85]. GPFHs would be groups of primary care doctors administering an independent budget, based on capitation funding, and used to pay for their own services and for the referrals of their patients to higher-levels of care.

Parallel to the above reforms, in countries such as the USA which traditionally relied more on private financing and provision of health care services, profound changes were also under way. First, a process of integration and increased co-ordination among independent units was taking place. Several indemnity insurers were evolving into new organisations, Preferred Provider Organisations and Health Maintenance Organisations, whose distinguishing feature was Managed Care. Managed Care is characterised by a much more pervasive control over providers' diagnostic and therapeutic decisions, mainly obtained through contractual discipline. The new contracts set prospective payments for providers (see below analysis of payment systems), quantity ceilings and utilisation reviews, and sometimes imply exclusive dealings, or other forms of vertical restriction. In other

---

[84] According to the theoretical model first proposed by the American economist Enthoven, providers would compete for health contracts, and GP Fund holders would compete alongside Health Authorities for public funding on the purchaser side. Competition was meant to promote a more efficient and consumer-responsive system, by offering: "greater rewards for those working in the NHS who successfully respond to local needs and preferences" ("Working for patients", 1989, pp.3-4). "Working for Patients" is the title of the White Paper outlining the reforms for the British National Health System.

[85] In 1997 the new Labour government initiated a reform plan replacing HAs and GPFHs with Primary Care Groups (PCG), later renamed Primary Care Trusts (PCT) on the purchaser side. PCT are commissioning and service provider agencies with an assigned pool of patients larger than that formerly administered by GPFHs, but generally smaller than HAs; they should promote a new coordination of community and primary care services. However, the key features of the 1991 internal market reforms are still preserved.

words, in these second group of countries a "spontaneous" evolution in the market (mainly driven by employers' need to achieve better control over escalating insurance costs) was creating the purchaser role (taken mainly by former indemnity insurers but also by groups of doctors), without an explicit intervention by government.[86]

Following the UK and the USA, during the '90s other European countries (such as Holland and Italy) and several other middle income countries of Latin America (such as Colombia), of Central and Eastern Europe, and of Asia implemented reforms inspired by the internal market model. The experience with these reforms in Colombia is presented in Box 4.1.

Case Study 4.1: Evidence from Colombia [87]

The 1993 Colombian health reforms strengthened the health revenue collection and pooling systems, they changed the criteria for allocating resources within the health system, and made health providers autonomous (hospital corporatisation). The core reform components were the following:

- On the resource mobilization front, the government raised the payroll tax rate on formal sector employees from 8 to 12 percent (the same rate holds for self-employed workers who report a salary above the national minimum), and hypothecated for health a fixed share of local governments' revenues (25% of the total local governments' revenue, according to Law 60 of 1993). Such revenue mainly consists of transfers from the central government, plus minor local taxes.

- The creation of a national contributory insurance scheme for formal sector employees, managed by Empresas Promotoras de Salud (EPS), and of a national subsidized insurance scheme, managed by Administradoras del Régimen Subsidiado (ARS). The EPS and the ARS receive a capitation funding for each beneficiary (la Unidad de Pago por Capitación – UPC), which is larger for children under 5, for the old and for women in fertile age, and they guarantee coverage of, or directly provide a prefixed package of health services to their beneficiaries. In prospect, the benefit package should become equal for all Colombians (Plan Obligatorio de Salud, POS), but so far it is more generous for those in the contributory regime than for those in the subsidized regime (approximately twice as expensive per capita, US$127 as opposed to $68 per capita in 1998).

- The creation of a national compensatory fund (Fondo de Solidaridad y Garantía, FOSYGA), entrusted with the role of supervising the financial flows in the system, of cross-subsidizing EPS in deficit with

---

[86] In fact, the USA government, under the Democratic Presidency of Clinton, drafted a comprehensive plan of reforms ("Health Security Act") inspired by the managed competition model first proposed by Enthoven (1985). The plan was proposed to Congress, and later on rejected by the Republican Party (see case study 4.4, in the text).

[87] This Case Study draws on a paper by Londono, Jaramillo and Uribe (1999).

resources from those in surplus (Subcuenta de Compensación), and of providing parts of the subsidies for ARS (the other part -69% in 1999- is provided by municipal governments) [88].

- The transformation of hospitals and other health providers into independent Prestadoras de Servicios de Salud (IPS), legally regulated as ESE (Empresas Sociales del Estado) and mainly funded by EPSs and ARSs according to the services provided to the insured population. Thus, direct subsidies to providers, previously assigned mainly by local governments, should have been progressively substituted by activity-based payments, according to the services rendered to the insured population. However, recently the central government stopped the process of phasing out of these subsidies.

The evidence available from Colombia after 1993 shows weaknesses and gaps in the reform process. Nonetheless, overall it shows a positive impact of the health reforms on equity and efficiency. In particular, some key macro-indicators of equity seem to have improved, in spite of the several difficulties encountered in the reform process, and in spite of socio-economic hardships the country continued to experience while the health reforms where in progress[89].

Looking first at the changes in social health insurance coverage, the evidence indicates that the Colombian health system before the reforms was extremely fragmented and inequitable, with marked differences in access to health services and their quality[90]. As a result of the weaknesses of the pre-payment schemes, the poorest decile of the population was spending on health a share of their income up to ten times higher than the richest decile, and approximately half of those who reported illness did not seek care because they were unable to pay[91].

After the reforms the number of people insured through both the contributory as well as the subsidized regime has sharply increased, raising the total number of people insured from 23% to roughly 60% of the total population. The improvement was particularly significant for households belonging to the poorest expenditure quintile in the country, who increased their health insurance coverage in the subsidized regime more than tenfold, and those belonging to the second quintile, who increased coverage approximately six times. In year 2000, of the total poor population in the country 35.3% was covered by the subsidized regime, 10.7% by the contributory regime, and 53.9% as yet did not have health coverage. Other major achievements of the reforms were the following: A) Between 1993 and 1997, health subsidies for the poorest quintiles of the population increased by 200%; and those for the second poorest quintiles by more than 100% (Sanchez and Nunez, 1999).

---

[88] The central government should also match FOSYGA's subsidies, a mechanism known as *paripassu'*. However, due to the difficult fiscal situation, such contribution was frozen.

[89] The growth rate in Colombia has been negative for most part of the last decade, unemployment raised from 7% in 1994 to 22% in 1999, and the civil war created hundred of thousands of displaced people.

[90] Social security covered health risks for approximately half of the workers in the formal sector; among these, only 20% enjoyed health coverage extended to their families. Overall, 72.2% of the adult working population in Colombia (and 90.8% of the poorer decile) was excluded from any social security benefits. Only 23% of the total population (approximately 7 million Colombians) enjoyed any form of health insurance coverage.

[91] Molina C.G. et al. (1993).

Meanwhile, health subsidies for the two quintiles of the most affluent sector of the population decreased 70% and 14% respectively. B) Disparities in the allocation of resources for health across regions and departments have been reduced with the process of decentralization and with the introduction of capitation-based resource allocation criteria[92]. C) Most importantly, utilization of services has increased across all income groups, as the following table shows:

Table 4.1: Colombia. Health Services Utilization Distribution, 1993 – 1997. [25]

	Ambulatory Consultations			Hospitalizations		
	1993	1997	Change %	1993	1997	Change %
Quintile 1	340,856	605,840	78%	176,045	205,982	17%
Quintile 2	436,223	827,147	90%	216,108	245,473	14%
Quintile 3	549,161	1,232,268	124%	201,606	322,482	60%
Quintile 4	590,331	1,384,333	135%	200,060	296,628	48%
Quintile 5	598,863	1,362,215	127%	179,550	333,724	86%
TOTAL	2,515,384	5,411,803	115%	973,369	1,404,289	44%

Source: Londono, Jaramillo and Uribe, 1999

Also the reform component which gave hospitals full autonomy had a significant impact (McPake et al. 2002). Before the reforms in Colombia efficiency indicators in the public sector were extremely poor. For example, occupancy rates in public facilities were lower than 40 %. As explained before, the reform programme included the creation of a purchaser-provider split and the transformation of public hospitals into 'autonomous state entities'. These were intended to contract with multiple competing insurers and the local health secretariat for the provision of services. A study aimed to track hospital performance in the post-reform period in Bogotá (McPake 2002). Trends in hospital inputs, production and productivity, technical quality, patient satisfaction and finances, and qualitative data based on interviews with hospital workers were collected. There was some evidence of increased activity and productivity and sustained quality despite declining staffing levels. The following two figures show trends, respectively in total admissions and in bed occupancy rates.

---

[92] The Corporation for Research and Development (CRD), and the Medellin Economic Faculty (CIDE) found that, between 1987 and 1995, regions that at the beginning of the period had relatively less public resources, witnessed a greater increase in funds available, and that as a result inequalities in per capita available by department have decreased. Bossert (2000) confirms the same results.

[25] Adapted from Sanchez F. y Nunez J. (1999). The numbers presented do not correspond to the total of the services offered in the country. They correspond to the estimations of the CASEN (1993) and of Quality of Life (1997) Surveys.

Figure 4.1: Trends in Total Admissions
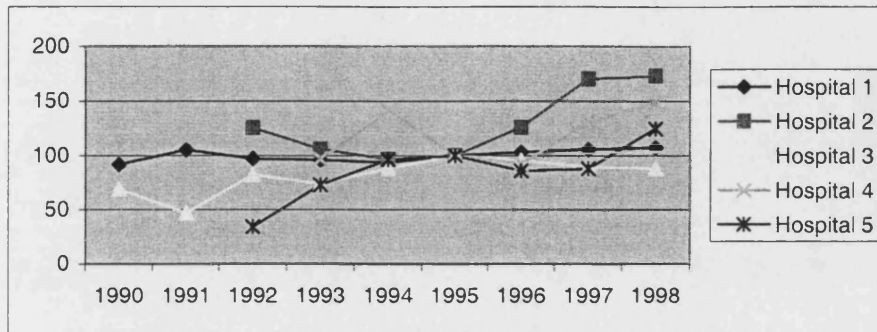
Source, McPake 2002

Figure 4.2: Trends in Bed Occupancy Rates
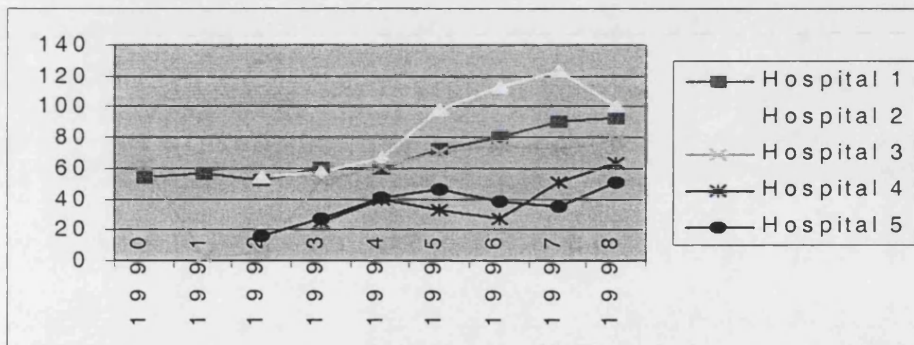
Source, McPake 2002

As the two figures above indicate, both number of admissions and bed-occupancy rates either remained stable or raised after the reforms. Qualitative evidence suggests that hospital workers have noticed considerable changes, which include greater responsiveness to patients but also a heavier administrative burden. Unfortunately, no data is available.

4.2 Resource allocation, purchasing and payment system (RAP) reforms

Resource allocation, purchasing and payment systems mechanisms define the criteria according to which funds, collected through different revenue sources and pooled together, flow within the health system, eventually reaching service providers[93]. Their place in relation to the other dimensions of health financing and service delivery can be visualized in Figure 4.3, taken from Preker et al. (2001).

Figure 4.3: Financial Flows within a Health System
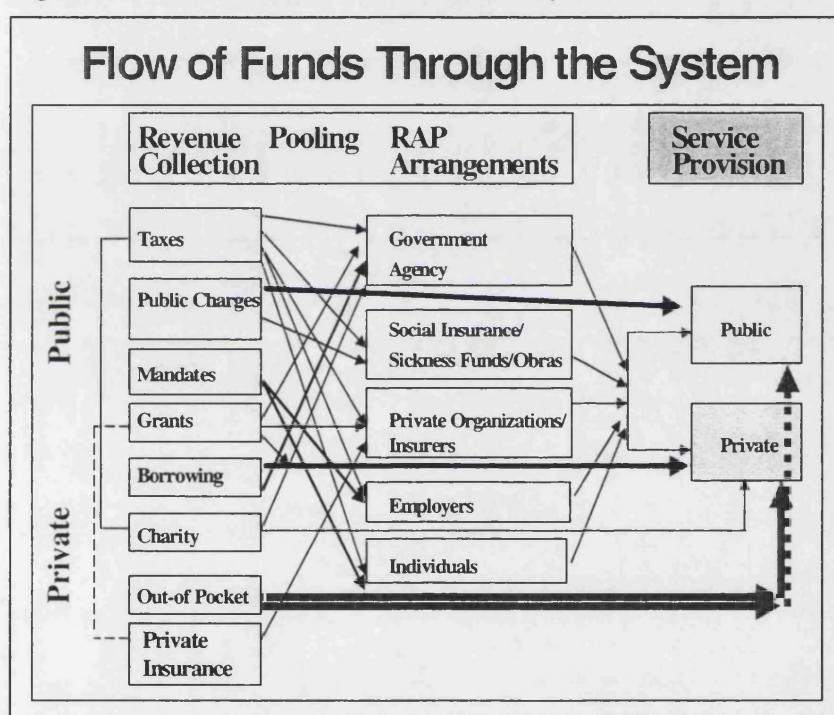
Source: Preker et al., 2001

RAP reforms, namely, changes in the criteria according to which funds flow within the health system eventually reaching service providers, were a key feature of the internal market or managed competition reforms described in the previous section.[94]

---

[93] Note that health financing includes both the alternative ways to fund health services, by collecting and pooling resources for health, as well as the criteria for distributing resources across purchasers and providers. RAP reforms are meant to change the latter, and not the former.

[94] However, note that internal market reforms frequently included wider changes, which also affected sources of revenue and the organization of service provision, as the Colombian case study just presented demonstrates.

This chapter will describe six key elements within RAP reforms. The first five intervene on the supply side, and the last one on the demand side:

- the interposition of a purchasing agency between patients and providers, entrusted with the role of commissioning/contracting services from (semi)independent providers;

- redesign of resource allocation criteria for purchasers, and of payment systems for providers, moving from input-based and retrospective towards prospective payment systems;

- redesign governments' priorities across services or levels of care, and limit public financing to a more restricted set of services (in several developing countries this component of RAP reforms has been identified with the introduction of the "Basic Benefit Package, or BBP);

- provision of explicit financial and other incentives/enablers to providers, linked with their ability to reach the poor/cure diseases that primarily affect the poor.

- financial incentives for patients/clients, such as vouchers, to stimulate consumption of specific health care services, such as prenatal care and institutional delivery.

All RAP health reforms implemented over the last twenty years in industrialized as well as middle income countries included a different combination of the above six elements. Table 4.2 synthetically illustrates their principal objective and content.

At the same time, some countries changed some specific RAP mechanisms, for example payment systems for providers, without affecting other dimensions of health financing or resource allocation.

Table 4.2: Different RAP Reform Components

RAP reform component	Priority setting	Resource allocation criteria	Purchasing/contracting	Providers' payment systems	Explicit incentives for providers to reach specific groups/cure certain diseases	Explicit incentives for patients to use specific services
RAP arrangement addressed	What services to subsidize and what to exclude?	What are the rules (formula) for transferring public funds?	To make or to buy decision	How to pay? How much to purchase and at what prices?	Whom to purchase for?	Whom to purchase for?
Purpose	Strategic definition of priorities for public financing and provision.	New criteria for allocating resources across regions, districts and purchasers.	Strategic purchasing; more transparency and results-orientation in public sector; engagement of the private sector in service provision.	Redefinition of the incentive regime for providers.	Reduction of exclusion of the poor from public services/sharper targeting	Increase utilization of essential services by the poor
Activities	Redefinition of priorities across services according to cost-effectiveness or other criteria.	From retrospective to prospective resource allocation.	Definition of contracts and service agreements, increased degree of market exposure for providers	Output/performance based payment systems for providers; capitation based funding of primary care physicians	Equity-motivated risk-adjusters to correct capitation funding. Explicit financial incentives for providers to cover the poor, or to provide care against diseases that affect the poor (ex. TB).	Specific enablers or incentives (such as vouchers) that contribute to treatment and travelling expenses
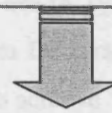
| Principal Objective | Max impact on disease burden, cost effectiveness, correction of market failures | Allocative and productive efficiency. Consumer-responsiveness. | | Equity, Financial protection | | |

4.3 Motivation of RAP Reforms

Several European Nations in the post World War II period provided universal health insurance coverage to their populations and created National Health Services to deliver uniform standards of care across different geographical areas (see Annex 2.1). The first NHS was created in the UK in 1948; then, during the 1950s, '60 and '70s several other West European Nations followed. The planning and delivery of most health services was organized through public, vertically integrated agencies. Most production decisions, such as volume and mix of inputs, salaries for health workers and other personnel, and capital investments, were taken by government at the central and or the local level. By these means, public health services effectively guaranteed an unprecedented degree of equality in access to health services, even if a certain degree of discrimination continued mainly because of cultural barriers and uneven education. National Health Services were also comparatively more effective in controlling total health expenditure.

However, over the 1980s this publicly financed, government-led, top-down approach to planning and delivery of health services was increasingly criticized for its inefficiency and poor results. According to these critiques, central planning, lack of accountability and of financial or other incentives linked to performance for health workers, created a rigid and non client-responsive delivery system. At the same time, the political and bureaucratic process, which determined the total amount spent and the share attributed to different services, hardly reflected consumers' preferences or any other economic rationale. Actual resources spent on health could not respond to the new expectations about standards and diversification of care that, during the last fifty years, accompanied the constant increase in income and education levels.

Over the post WWII period several middle and lower income countries established and staffed an extensive network of health facilities at the primary, secondary and tertiary level. Given their colonial inheritance and socialist orientation, several of them tried model their health system organization on the basis of West European National Health Services. The Alma Ata Declaration, subscribed in 1978 by all Nations but a few, marked the high point in this struggle to achieve "free health care for all". In reality, MIC and LIC did not and do not have the tax base and public resources that could finance the operation and functioning of a universal public delivery system. Instead of following the European path towards publicly financed, publicly provided health care systems, because of tight fiscal and other political economic constraints, LIC with a few exceptions (see Box 4.1) ended up with a (mainly) public, but (mainly) privately financed delivery system. In

rural areas, unqualified or semi-qualified private providers continued to be the first contact for the majority of poor households, unable to afford formal private or public services. Over time, a private delivery sector –for profit and not-for-profit- also grew in these countries. The private sector entered in a sort of awkward symbiosis with the public delivery system, where the private sector benefits from the disorganization and the lack of financial resources in the public sector, and where the public sector can survive unchanged because the frustrations its imbalances continuously create (among patients, as well as physicians and nurses) are relaxed by the possibility of exercising/seeking private services. In this situation, doctors and other health employees, supposedly employed full-time in the public sector, increasingly earn most of their income in the private sector, and use their public employment mainly to gain a clientele for their private activity and to utilize expensive equipment they could not otherwise afford.

In summary, in the majority of these poorer countries, a combination of government and market failures more severe than in industrialized Nations have resulted in poor spending choices, distorted allocation or resources, poor quality of services, unequal distribution of resources for health care in favour of the rich (see Chapter 6), etc. The government health care system has been incapable of universally providing at acceptable standards even essential services such as maternal and child care. A number of studies has found a high level of absenteeism in government facilities (see for example, results from study in India presented in WDR 2004, The World Bank: WDR, 2004, Chapter 2, Table 1.2b), widespread dual job holding by doctors, and that "bypassing" of low-quality services offered by public facilities is a common phenomenon (Akin and Hutchinson, 1999. For Kenya, see Mwabo, 1993; for the Dominican Republic, Lewis et al., 1992, and for Pakistan, PIEDR, 1994). The high price-elasticity of demand for government health services, particularly observed for child services and among the poor (for Uganda, see Giusti, 2000[95] and for Indonesia, Gertler and Molyneaux, 1990) may conceal a low evaluation of these services.

Box 4.1: The relationship between government health expenditure, level of income, and health outcomes

A stream of literature has recently questioned the effectiveness of publicly-financed and provided services, particularly curative services, in achieving better health outcomes. In a multi-country regression, Pritchett et al, 2001, found that the level of health expenditure and health inputs officially recorded, such as the number of

---

[95] In 1997 a natural experiment was recorded in Uganda. Non-profit facilities raised their user-fees in August (up to roughly US $ 1.5 per child outpatient visit, and up to US $ 7.5 per adult outpatient visit) and then lowered them again in November, after the government committed to subsidize them. Giusti (2000) recorded extremely wide variation in service use, especially for childcare.

doctors or number of health facilities, produce a negligible impact on priority health outcomes). By contrast, other evidence (see for example Wagstaff, 2004) indicates that in countries characterised by better governance, there is link between health expenditure and health outcomes. The evidence also indicates that real, as opposed to "paper", health inputs and service quality do produce a strong impact on utilization and on outcomes (Alderman and Lavy, 1996). For instance, Lavy et al. (1996) show that in Ghana drug availability and weekly hours of availability of physicians produce a strong impact on child survival, on child height and weight for height.

These studies also question the effectiveness of the traditional ways of channelling foreign assistance in the health sector to developing countries, mainly focused on the expansion of the infrastructure base of the government health sector. While these continuous investments have made a tangible difference to physical infrastructure and delivery capacity, the outcomes in terms of actual improvement in services – particularly to the poor and disadvantaged groups – have not been as obvious. [96]

By looking at the variations across countries and at the few success stories available, it is also becoming recognizable that the results obtained by a few countries in terms of health outcomes and in terms of quality and accessibility of essential health services are achievable also at relatively low levels of economic development. For example, the international evidence indicates a strong relationship between average national income (measured by per capita GDP) to priority health outcomes, but it also shows that this relationship is by no means universal or automatic because a few low income countries perform exceedingly well in terms of health outcome. Among the good performers given the level of socio-economic development, we find Cuba, Sri Lanka, the Indian State of Kerala, and Costa Rica. For instance in 1997 Cuba and Costa Rica had a per capita Gross Domestic Product respectively equal to $3,100 and $3,810, an under-five mortality equal to 9 and 13 respectively, and a life expectancy equal to 75.7, 76.0 years, which are indicators comparable to those of the wealthiest countries of West Europe and North America (World Bank, 1999). In 1998 life expectancy in Sri Lanka was 73 years, and the infant mortality rate was 15.4 per 1,000 live births; in 2001 income (GDP) per capita was only US$837. By comparison, in Brazil GDP per capita was $2,940, but 20 percent of children were still reported as undernourished, the infant mortality rate was 50 per 1,000 live births, and life expectancy was about 67 years (World Bank 1999).

In conclusion, the evidence indicates that public services, particularly public health services, can potentially make a huge difference for the poor. In a study based on a sample of 35 countries (where nationally representative household surveys were conducted), Bidani and Ravallion (1997) show that public health spending matters for the poor (measured as those below US$2 or US$1 per day at 1985 purchasing power

---

[96] A similar point is stated in the 2001 HNP Chapter for the PRSP Sourcebook: "One point needs emphasizing, namely that funds linked to PRSPs –whether debt relief or IDA credits- will have a far greater impact on poor countries' health levels if they are accompanied by a thorough review of existing policies, and by a willingness to link new spending with reforms that make health systems work better, especially for the people they tend to serve less well –the poor (2001 HNP Chapter for PRSP Sourcebook, p.5).

parity), although its impact is insignificant on the richer segments of the population and on the population as a whole. More recently, Gupta et al. (2001) reach similar results. These results are not surprising, if we consider that the poor are less able to substitute private care for public care, when the latter is lacking or is of extreme poor quality (see Chapter 6, section 6.2).

Against the background described above, RAP reforms can be easily justified: they are an attempt to respond to the above issues of efficiency, quality, and equity in the public as well as the private sector, and increase the "value for money" that patients obtain from the health system. Their premise is the recognition that "incentives apply to governments and not just to markets" (Stiglitz, 1999, pp.3-4). Hence, some RAP reform component try to replicate within the public sector some market-type incentives also for public providers, in order to grasp some of the advantages of markets vis-à-vis government provision in terms of flexibility and efficiency (see below, Table 4.2 from Preker and Harding, 2003), but at the same time to respond to the accessibility and quality issues that affect private markets.

Table 4.3 Comparative advantages and disadvantages of the public and private health sectors

Issue	Public Sector	Private Sector
Equity and access	• Targets services for poor and vulnerable populations • Attentive to geographic disparities	• Individuals who can pay favoured; the poor and vulnerable excluded or ignored • Services concentrated in population centres
Public health, preventive and curative care	• Emphasis on preventive and public health services (public goods with large externalities) • Extensive system of hospitals and curative care centres often maintained	• Little attention to preventive and public health services without special incentives • Emphasis on curative care services valued by paying customers (private goods)
Management	• General dependence on political and legislated direction • Difficulty recruiting qualified managers • Hierarchical bureaucracy with diffused accountability • Commitment to public service compromised by vested personal interests • Restrictive range of discretionary authority, less flexibility, less innovation	• Greater reliance on information for decision-making and planning • Recruitment of managers limited primarily by cost/benefit considerations • Smaller and more focused authority structures • Greater synergy between business and personal interests • Broader range of discretionary authority, greater flexibility, more innovation

Issue	Public Sector	Private Sector
Customer orientation	• Heterogeneous constituency with wide range of expectations • Limited attention to customer convenience and comfort • Indirect accountability for customer satisfaction	• Focuses on relatively narrow range of customer needs and wants • More attentive to customer convenience and comfort • More direct accountability for customer satisfaction • May exclude poorest and sickest
Flexibility	• Extensive infrastructure of owned facilities • Restraint of civil service system • Slow to respond to changing market conditions because of political and budgetary commitments to ongoing programs	• Adaptable access to infrastructure through rentals and leasing • Flexible employment and pay practices • Quicker response to changing market conditions
Financing	• Access to tax revenues • Weak incentives to be cost from sales or contracts conscious or cost efficient • Programs financed primarily through historic budgetary allocations • Limited access to private capital or donations	• Dependent on revenue flows • Attentive to cost and price • Needed but unprofitable services possibly curtailed or discontinued • Resources assigned to profit centres • Sensitive to cross-subsidization and cost shifting • Access to capital markets or donations
Competition	• Possible monopoly on selected services reinforced by regulation and subsidization	• Subject to competitive pressures from public and private providers • When entering the market, interested in increasing contestability and, once successful, interested in decreasing contestability (to restrain competitors)

Source: Harding and Preker (2003), The World Bank.

Note that RAP reforms have not been justified because they intended to improve services specifically for the poor. However, the above discussion suggests that any health reform that succeeds in improving quality and effectiveness of health services in general, without reducing accessibility, it is also likely to have *ipso facto* a positive impact on the poor (see Chapter 6), because it is the poor who mostly have to rely on the government for financing their health expenditure (see Chapter 6, section 6.2).

4.4 Main RAP Reform Components

The following sections will briefly describe the six main components of RAP reforms introduced in section.

4.4.1 Resource Allocation Criteria

'Resource allocation' (RA) criteria define the flow of financial resources from the centre to decentralized levels of government (such as regions and districts) and to health purchasers. RA criteria are closely related to payment systems, a term usually utilized to indicate reimbursement criteria for service providers. Over the last two decades, several of the industrialized countries implemented resource allocation (RA) reforms in the health sector, as well as other sectors such as education. They shifted from mainly retrospective criteria (based on historical allocations), towards population-based, prospective RA formulas for financing local levels of government and health purchasers. A resource allocation (or payment system) is referred to as retrospective, when the cost of the activities implemented (in terms of volume of inputs and unit cost of service) are completely reimbursed ex-post, and prospective when the amount given is determined ex-ante (for example, according to the population covered), and completely independent of the actual costs incurred.

4.4.2 The Priority Setting Component of RAP Reforms

During the decade of the '90s several countries have introduced explicit rationing or priority setting policies, based on the definition of an essential package of health services (BBP, Basic Benefit Package, or ESP, Essential Services' Package), whereby government financing is limited to a subset (in terms of services or people covered) of all the health services demanded by the population. In some cases, rationing/reprioritization policies have been implemented at the sub national level (for example in Oregon, USA). In other case, such as Uganda, reprioritization policies defined at the central level have clashed with the concomitant decentralization process[97]. In low-income countries (LIC) the definition of the services to be included in the BBP was heavily influenced by the 1993 World Development Report (World Bank, 1994).

---

[97] In the case of Uganda, for example, the amount of public funds spent on primary health care at the district level declined dramatically following decentralization despite it being a clearly stated national priority. The response there was to shift back towards a direct allocation approach in the short term by earmarking funds to the health sector. This has caused problems, as some saw it as inappropriate decentralization. They thought that the new RA processes, by allocating a block grant to the local level, was supporting capacity building and more effective management at district level. Others saw it as a legitimate move by Government to ensure national priorities were followed. Recently in Uganda the process of decentralization has been moving ahead again. The role of district hospitals in delivering primary health care services has now been recognized and attempts are being made to allow these units to qualify for Poverty Action Fund financing.

The content of an essential service package is obviously extremely important. Ensuring equal access to an inappropriate range of services does not get us very far – for example a pro poor policy requires a greater focus on those diseases which disproportionately affect the poor, and on those services that are likely to lead to the greatest health improvements for the poor. As we discussed in Chapter 2, the two main principles which should guide the priority setting process (the process of deciding what is 'essential' and what is not should be: a) correcting market failure; and b) improving the pro poor impact of health policies. Annex 4.1 discusses to what extent these two principles have guided the design of reprioritization policies. Moreover, the theoretical discussion leads to the conclusion that priorities should be defined as packages of services across levels of care, rather than as particular levels of care; however, because of administrative and political economic constraints, where reprioritization policies have concretely been implemented they have mainly consisted in the redistribution of resources across levels of care (away from tertiary and towards primary care). For example, in Bangladesh, the Essential Service Package has been defined and measured as: 'all primary care interventions delivered at *thana* (district) levels and below'.[98]

4.4.3 Purchasing/Contracting

The interposition of a purchasing agency between patients and providers is a relatively new phenomenon. Until two decades ago, in every health system resource allocation decisions were organized either from the central government, as in National Health Systems, or directly through private patients-providers interaction, with little control over them by third-party insurers.[99] "Purchasing" reforms were introduced first in some industrialized countries, such as the UK (1990), and then during the '90s in several middle income countries. As explained before, all of these countries entrusted newly formed agencies (such as social insurance funds, or groups of mainly primary care doctors), or existing ones (such as local authorities, or private insurers in the USA and Latin America) with the purchasing role. Purchasers discipline their relationship with health providers through service agreements or "contracts". Here the term "contract" is used loosely, referring both to the situation where there is a formal legal agreement between two private and

---

[98] Such definition excludes ESP services being delivered in hospitals, and it assumes that all spending at *thana* and below is on essential services (Bangladesh Public Expenditure Review 1999/2000).

[99] For instance, neither the British nor the American health systems, which exemplified the two opposite mechanisms for allocating resources among industrialized countries, had agencies entrusted with the purchasing role. The British National Health System was, at least formally, a centralized and vertically integrated system where Health Authorities, the public agencies locally responsible for planning and provision of health services, directly managed hospitals. By contrast, in the American health system insurers reimbursed health providers mainly on a retrospective, fee-for-service basis (see discussion in the text).

totally independent institutions (as it is normally the case in the USA health system, or other health systems where purchasers are private agencies), as well as where the purchaser and/or the provider are public agencies under the umbrella of a National Health Service or a Social Insurance Fund (as in several European countries).

4.4.4 Payment system (PS) reforms

Several countries of Europe, Latin America, Asia, and Africa implemented PS' reform over the last decade. The general trend has been to move away from input-based PS, towards activity, case-based payment systems, such as DRG-based payments for hospitals, and capitation payments for primary care doctors. With the new resource allocation criteria introduced through RAP reforms, part of the financial risks for unexpected increases in unit costs and volumes of service is shifted from the central government on to health purchasers, and, in turn, to providers through prospective payment systems. We present the major PS and the impact of PS reforms in Chapter 5.

4.4.5 Supply-side Incentives in Pro-Poor Health Programmes

Some countries have introduced explicit performance-related financial incentives, particularly in their vertical programs against communicable diseases, to achieve better coverage and improved health outcomes among the poor. For example, explicit financial incentives to improve immunization coverage was introduced under GAVI, the global vaccine initiative, where an amount equal to US$ 20 is assigned for each additional child immunized (DP3) in countries with immunization coverage rates comprised between 50 and 80 percent. Similar explicit financial incentives, to be given directly to providers, are being considered under the Global Fund to fight malaria, TB and HIV/AIDS.

4.4.6 Demand-side Incentives in Pro-Poor Health Programmes

Explicit demand-side incentives can also and have been utilized to stimulate consumption of specific health and reproductive health services. Their rationale is to stimulate demand for health services, which is usually lower among the populations needing them most (see Annex 2.2), particularly services with large positive externalities (such as vaccinations, maternal and child care, communicable disease treatment; see Chapter 2). To stimulate the consumption of these services by

the poor, specific enablers or incentives are given (such as vouchers that contribute to treatment and travelling expenses) that lower the opportunity cost of receiving care (cost that includes time lost for gainful employment, travelling, etc.).

4.5 Trade-offs in RAP Reforms

In principle, the optimal RAP arrangement should be easy to find. Simply, it is the one which would induce providers to perform high quality, cost-effective treatments in an efficient way, while at the same time promoting an equitable allocation of resources within the health sector. In reality, such optimal RAP arrangement does not exist, as we shall see in this and the following chapter, trade-offs among potentially equally desirable aims occur. Each reform component is appropriate to achieve only one of those aims, and, by doing so, in general, it moves the health system further away from achieving other equally desirable aims.

Note that the health policy literature does not, on the whole, talk in the language of trade-offs. This has the (probably unintended) implication that it is possible to achieve all objectives simultaneously, if only the right policy intervention, being it a specific resource mobilization or resource allocation mechanism, or a specific payment system for providers, can be found. In contrast, we argue that in general each health reform component comes with potential negative "side" effects. Given the existence of trade-offs, health reforms should contemplate a plurality of interventions (in resource mobilization and allocation, payment system, management, etc.), able to reinforce each other and at the same time to correct the potential negative impact of each intervention taken in isolation.

In this section, we introduce a conceptual framework which will be utilized to illustrate the potential trade-off between equity and efficiency. Let us utilize a graphical illustration of the (expected) Health Frontier (HF)[100] for two individuals, Mr. Rich (R), and Mr. Poor (P):

---

[100] The Pareto Frontier is the set of all Pareto optimal allocations. Each point on the health Pareto Frontier for Mr. Poor and Mr. Rich is found by fixing the level of expected health of one of the two, and by maximizing expected health for the other.

Figure 4.4: RAP reforms. Moving Towards the Health Frontier and/or Improving the Health of the Poor?
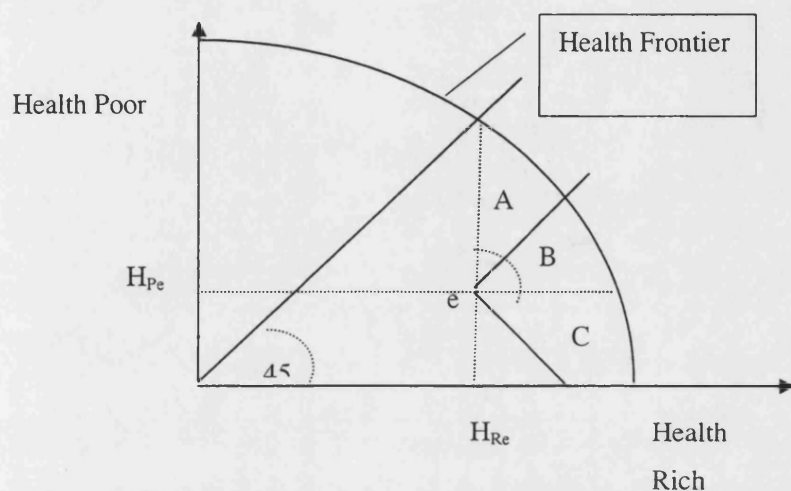
The above figure simplifies a situation where the achievable health gains are constrained by the available amount of resources, and where the choices on the use of the existing health resources may create an equity-efficiency trade-off. Each point in the graph represents the (expected) health, or health gains for the two individuals associated with a different allocation of health resources and a different incentive structure within the health system. The down-sloping Health Frontier (HPF) curve represents the set of optimal points, where it is not possible to further increase expected health gains for one of the two individuals, without decreasing them for the other.

Suppose that in the initial situation the health sector is characterized by a suboptimal equilibrium (point e in the graph), where, due to a distorted allocation of resources or lack of incentives, the system is not attaining the maximum health gains for the two individuals. Graphically, the allocation in correspondence to point e is not optimal, being an allocation internal to the HPF.

Suppose that:
   a. efficiency is measured by the sum of the health gains for the two individuals[101], and
   b. equity is measured by the difference in the total health enjoyed by the two individuals (egalitarian view, see Chapter 6).

---

[101] Note that the concept of efficiency used here is different from that of Pareto-efficiency. It is possible to increase the sum of the health enjoyed by the two individuals also by moving towards points where one of the two individuals is worsening her/his health status.

Starting from point e, graphically one could represent an increase in the level of efficiency (for example, because a new capitation-based payment system for providers is introduced; see next Chapter) as a movement towards the curve HF. However, as the graph indicates, there are three different ways by which one can move towards the HF. The first is to move from point e in area C. Note that this would not represent a Pareto improvement (see Chapter 2), because Mr. Poor is adversely affected by the change (although losing less than what Mr. Rich gains). It also leads to worsen equity, according to the above measure, as the health differential between the two individuals widens. The second and third ways are to move from point e in areas A or B. Note that this would represent a Pareto-improvement, as both individuals are experiencing positive health gains from e to any point in A and B. However, improving efficiency by moving to A or B is not the same. Starting from e, in B the health differential between the two individuals widens (and so equity worsens), whereas in area A inequality is also reduced.

In conclusion, starting from any inefficient initial equilibrium, we can point at the existence of an equity-efficiency trade-off any time as a consequence of health reforms health outcomes improve overall but also health disparities widen.[102]

4.5.1 Trade-offs in RAP Reforms: Purchasing

In this section, we introduce a first example of the trade-offs associated with RAP reforms, focusing on the purchasing/contracting component. The next chapter, Chapter 5, will investigate more in depth the different trade-offs associated with the payments systems component.

In purchasing/contracting, we shall describe the following trade-offs:

1. Between the aim of avoiding adverse selection, which can be achieved by having purchasers with large risk-pools, and the aim of allowing greater choice for clients and consumer-responsiveness, which can be achieved by stimulating competition among purchasers, and containing their size and market share.

2. In contracting, the trade-off between the aim of maintaining uniform standards of care everywhere, which requires central control, and the aim of enhancing results-orientation and performance, which require more competition and incentives at the individual level.

---

[102] For a critical review of the different concepts and measures of equity currently used, see the section on equity in the text. In Figure 4.5, note that if we were only interested in the absolute health of the poor (as argued by Rawls, 1971), and not in overall disparities, movements towards points in area B will be judged

Purchasers: Consumer Choice and Client-Responsiveness versus no Risk Selection

Under the new RAP arrangements, a trade-off arises between the aims of limiting risk selection and of pursuing a positive redistribution across different risk-types through the pooling mechanisms, and those of increasing purchasers' consumer-responsiveness and patients' choice.

Reformers would like purchasers to reflect patients' legitimate preferences. For example, to promote through contracts 'good' providers and to penalize 'bad' ones, according to patients' perceived quality. Regardless of all other possible forms of regulation and control, the only true guarantee that purchasers will take into account patients' preferences in their choices is to allow patients to leave them when they are dissatisfied (to allow them to vote with their feet), and to financially penalize those purchasers that are not able to attract patients. A precondition for this to happen is that patients be free to choose among several different purchasers. Otherwise, providers would most likely be able to collude with purchasers against patients' interest, and unsatisfied patients would have no alternative. Following these principles, several RAP reform proposals (for instance, Poland and Czech Republic among former socialist countries, and Mexico, Chile and Colombia in Latin America) intended to create and safeguard competition among purchasers.

At the same time, at least in health systems in principle committed to universal insurance, reformers also require that purchasers continue to guarantee coverage at equal standards, and not to select patients, according to their health risk or income.

The two above aims, to safeguard "good" purchasing through competition and to promote equitable and efficient pooling, are intrinsically in conflict with each other. The reason is because the same issues that plague private health insurance markets, which were investigated in depth in Chapter2, resurge for purchasers if/when they are allowed to compete with each other. The "market" or "quasi-market" equilibrium may well be suboptimal[103], whereby "low or good risks" might receive incomplete coverage and "high or bad" risk individuals may not be covered. "High" risk usually

---

positively from also an equity perspective. See also Case Study 6.2, based on Waters (2000).

[103] The theory of insurance markets with adverse selection proved that some positive level of redistribution in general characterizes second best equilibria across different risk groups. The free market equilibria may fail to achieve such cross-subsidization, because, in Rothschild and Stiglitz's (1997, p. 74) words, "One of the natural methods that insurance companies use to compete, underwriting, destroys or considerably limits what

coincide with the old, or chronically ill individuals, and so competition and market segmentation may lead to under coverage or outright exclusion from each purchaser's pool of the more vulnerable segments of society. [104]

As far as this trade-off is concerned, the 'size' of the purchaser (measured by the number of people covered) is a critical dimension. As the size of the pooled population increases, the degree of consumer-responsiveness decreases, because competitive pressures become weaker, and because purchasers become more distant from patients. However, also the incentive to risk select potential customers diminishes, as the purchaser can exploit substantial economies of scale in risk pooling, and cross-subsidies within their pool become easier to sustain (see Chapter 2, section 2.6). In addition, the larger is the risk-pool, the lower is the variability in individual income, and so the more efficient it is, due to a very simple statistical reason that we explain in Box 4.2.

As far as risk-selection is concerned, in health matters are made worse by the extreme concentration of risks.[105] For small purchasers, who receive a capitation-based Resource Allocation[106], have a disproportionate share of high risks, and are unable to dilute them in a large pool of insurees, risk selection may become the only financially viable strategy.

In theory, as we mentioned in the conclusion of Chapter 2, the resource allocation criteria can provide a strong incentive not to risk-select, by providing appropriate rewards to purchasers who recruit and retain high risks (risk adjusters[107]). Insurers/purchasers that have a disproportional share

---

insurance markets can do".

[104] Note that this problem cannot be solved simply by assigning the pooling and the purchasing function to two different agents. Even if resources are collected and pooled "upstream" (at the central level), prospective resource allocation and payment systems shift all or part of the financial risk "downstream" on to purchasers or the lower-level agent who is paid prospectively (see discussion on payment systems in Chapter 5).

[105] Evidence from several countries shows that in each country approximately five/seven percent of the population is responsible for roughly two-thirds/three quarters of total health expenditure

[106] Rice and Smith (2000) write: "A number of methods can be envisaged for funding the purchasers –which we have defined in the text as resource allocation mechanisms-. They can be summarized under three headings:

- (a) full retrospective reimbursement for all expenditure incurred;

- (b) reimbursement for all activity on the basis of a fixed schedule of fees (using, for example a system of diagnosis related groups);

- (c) prospective funding on the basis of *expected* future expenditure, using fixed budgets.

These three forms of resource allocation imply a progressive shift of risk from the national funder towards the health care plan. The trade-off described in the text is relevant if and only if (b) or (c) are in place, which is the situation in all countries that have implemented RAP reforms. See also discussion of payment systems for providers, in Chapter 5.

[107] Note that the 'risk" adjusters that serve redistributive aims are conceptually different from those that are meant to limit risk-selection. The former can also be based on geographic and social variables, while the latter are based on individual data. However, the use of both types of risk-adjusters becomes more difficult and less

of chronically ill people, for instance, could receive a premium support or subsidy from a compensatory fund, which would make these clients more attractive to them. However, existing risk adjusters can predict only a minor part of service utilization and expenditure in the future, and therefore they cannot completely eliminate the incentive to risk-select.

Another way to prevent risk-selection would be to allow small purchasers to reinsure with larger ones, but then their incentive to curb expenditure would also be diluted accordingly. The point is that in order to limit risk-selection, it is necessary to pay a price either in terms of less competition, less consumer-responsiveness, or in terms of less tight incentives for cost control (see also next chapter on payment systems for health providers).

Box 4.2: Risk-pooling and the size of the purchaser

Risk pooling can be represented as a mechanism through which risk-adverse individuals insure against their risk of income variations, by pooling together their resources. Consider n individuals, and let each individual's income be a stochastic variable y_i , where i=1,...,n, characterized by a certain statistical distribution, with expected value equal to μ_i and variance equal to σ_i. For simplicity, assume that the stochastic variables y_i are independent and that the expected value and the variance are equal for all individuals ($\mu_i=\mu$, var $y_i =\sigma^2$, for each i). Consider what happens if in each period individuals pool their incomes into a mutual insurance fund, that then assigns to each one the mean of the aggregate income received, thus compensating the variations in individual incomes occurring over time. The variance of the aggregate income received by the fund is equal to n times the variance of individual incomes (the variance of the sum of independent stochastic variables is equal to the sum of the individual variances):

$$\mathrm{var}(y) = \mathrm{var}\left(\sum_{i=1}^{n} y_1\right) = \sum_{i=1}^{n} \mathrm{var}(y_i) = n\sigma^2$$

However, after the individuals pool their income in the insurance fund, the variance of individual income is equal to the variance of mean income. This is equal to [108]:

$$\mathrm{var}(y_i) = \mathrm{var}\left(\sum_{i=1}^{n} \frac{y_i}{n}\right) = n\,\mathrm{var}\left(\frac{y_i}{n}\right) = n\frac{1}{n^2}\mathrm{var}(y_i) = \frac{n}{n^2}\sigma^2 = \frac{\sigma^2}{n}$$

---

likely to limit negative side effects of competition, when purchasers are smaller and when competition among them is more intense.

[108] In the following equation, we refer to a very simple propriety of variance, namely that:

$var(ax+b) = a^2\,var\,x$.

Thus, by pooling their income together through a risk-pooling mechanism, individuals are able to obtain income equal to the mean income, with a variance equal to (1/n) of individual variance. It is clear that as the number of people in the pool increases, the variance of the mean income shrinks approximating to a situation of certainty (Central Limit Theorem). Also, note that the "risk-pooling" function is conceptually distinct from the redistributive function that may also be achieved by pooling funds. In our example, redistribution of income across individuals may be achieved if ex-ante different individuals had different expected income. By joining the pool the rich would then trade an expected higher income with greater variability with a lower expected income with lower variability. Again, ceteris paribus increasing the size of the pool would increase the convenience of pooling.

In conclusion, in the choice of the purchaser's size (number of patients or maximal number of patients that the purchaser can "cover") a balance between the terms of the consumer-responsiveness, effective risk-pooling trade-off must be found. The optimal size of purchasers will depend on the risk distribution of the population, on how accountable even large purchasers can be made to pressures from patients, and on the existing possibilities of risk adjusting or re-insurance. Each country must find its own "optimal point" along the lines of this trade-off. We briefly discuss in the Case Study that follows the debate that took place during the '90s in the US and the UK concerning the optimal size of the purchasers, in light of the trade-off just explained.

Case study 4.2: Purchasers in the USA and the UK

We can represent in a graph the point chosen by the different purchasing reforms proposed in the UK and the USA at the beginning of the 1990s, along the terms of the above trade-off:

Figure 4.5: Purchasers: Trade-off Between Consumer-Responsiveness and Effective Risk Pooling
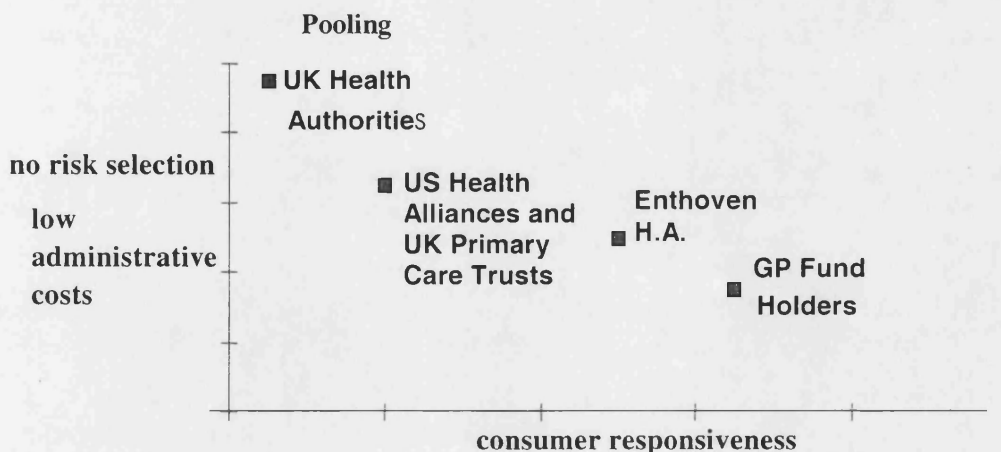
In the US, new institutional purchasers (health alliances) were envisaged by the Health Security Act (the Democratic Administration's plan of reform of the health system presented in 1993). The policy debate centred

on the role and "size" which health alliances would take. The first issue was the relationship of the new agencies vis-à-vis health insurers. The plan envisaged that health alliances would initially take the role of purchasers ("demand sponsors"), but not replace insurers. However, it was intended that over time a process of integration between insurers and providers in health care plans would progressively eliminate from the market traditional 'passive' insurers. Second, the "size" of the population covered by each health alliance, and their mutual relationship (competition, versus territorial demarcation of each one's area of operation) was debated. Finally, the Health Security Act came out with a proposal for health alliances which was extremely different from that originally planned by Enthoven, the health economist who first inspired the managed competition model and suggested interposing purchasing agencies between patients and providers. It is interesting to briefly describe that debate in the light of the trade-off explained above. Enthoven (1985, 1991) envisaged a system strongly based on competition both on the provider as well as the purchaser side. Therefore, he was in favour of the abolition of tax-exemptions in order to enhance price-sensitiveness of consumers, and wanted small health alliances in direct competition with each other. By contrast, the Health Security Act, ultimately preferred to emphasize the aim of universal coverage, and planned large health alliances to maximize economies of scale in risk pooling (competition for the market was left, but not competition within each market). Note that, also because of the strong opposition to the Health Security Act by private insurers, the Act was abandoned when the Democratic Party was defeated in the 1994 Senate elections. In terms of the above trade-off and relatively to the European experience, the American system continues to favour consumer-responsiveness of purchasers and does not really prevent risk-selection. Risk-selection is only partially mitigated as insurers are allowed to fix premiums according to experience rating (which, however, makes the American health system perform very poorly in terms of "equity of access"; see Chapter 6), and by direct government intervention through the programs Medicare and Medicaid which offer basic coverage to some of those excluded from private coverage.

In the case of the UK, the 1991 quasi-market reforms were introduced within a system that intended to preserve universal coverage of the population. The purchasing role was assigned to District Health Authorities and GP Fund holding schemes. The former were able to take full advantage of large risk pooling, while they largely remained unaccountable to the patients they served. By contrast, GP fund holders were much smaller in terms of population covered, could be abandoned by unsatisfied patients, and actively compete to find better conditions of service for their clients. However, their ability to manage risk was never really tested (Matsaganis and Glennester, 1994), because they always had the possibility to shift more expensive patients to HAs, and to game the system (for example, by recurring to emergency admissions).

In 1997 the new Labour government initiated a reform plan replacing HAs and GPFHs with Primary Care Groups (PCG), later renamed Primary Care Trusts (PCT) on the purchaser side. PCT are commissioning and service provider agencies with an assigned pool of patients larger than that formerly administered by GPFHs, but smaller than HAs; they should promote a new coordination of community and primary care services. The key feature of the 1991 internal market reforms, the purchaser-provider split, is still preserved.

One final observation on this first trade-off: as purchasers integrate downstream and start directly offering a wide range of health services (as in the case of American Health Maintenance Organizations, or of health purchasers in several countries of Latin America), the incentive for risk selection can be limited, because cost reductions may be achieved more effectively by intervening in the process of service provision instead of selecting less risky and costly patients. However, the drawback of this solution is that it may cause restrictions in the availability and the quality of treatment, as the trade-offs concerning payment systems indicate (see Chapter 5).

Efficiency versus the Right to Equal Treatment

The second trade-off we discuss is between the objective of promoting more productive efficiency through competition and contracting out of services, and the objective of maintaining similar health service standards everywhere.

The RAP health reforms initiated in many countries are based on the view that one of the more important factors explaining providers' inefficiency and scarce consumer-responsiveness is the lack of competition. Thus, as we wrote before, wherever they have been introduced, RAP reforms have always emphasized the need of more competition.

However, as competitive pressures in the health care delivery market grow, phenomena of horizontal and vertical segmentation similar to those described in Chapter 3 are likely to arise. As we saw there, all service markets are characterized by the fact that product quality is very important and preferences are heterogeneous: therefore, providers will try to relax the competitive pressure, by differentiating their services from those of their potential competitors. Horizontal (variety) differentiation, such as geographical differentiation, is exactly what reformers seem to intend: by horizontally differentiating their services, providers specialize in those services where they enjoy a comparative advantage in terms of knowledge of market conditions or technology. As a result, patients are given wider choices, and efficiency is enhanced.[109]

---

[109] Intense horizontal segmentation also brings the less desirable result that price competition is reduced (as each provider concentrates on a particular market niche), and therefore that the endogenous pressure for cost containment could be quite modest in the long term. However, incentives for cost containment may still be introduced by appropriately designed payment systems and contracts between purchasers and providers, as we shall discuss in the next chapter.

By contrast, vertical (quality) segmentation is not at all the result that RAP reforms, particularly in countries committed to universal coverage at similar standards of care, would like to see. By vertically differentiating their products, providers attack market sections characterized by different ability to pay, focusing either on low quality, low cost (and low price) services, or on high quality, high cost (and higher price) services (see Chapter 3). In the presence of a public provider bound to universal coverage operating alongside private competitors allowed to serve those who are able to pay directly or through private insurance coverage, vertical segmentation can degenerate in cream skimming, whereby private providers attach only the profitable part ('the cream') of the market, leaving the unprofitable part ('skimmed milk') to the residual public provider. In the health sector, the "cream" could be constituted by elective services, or in terms of individuals, they are identifiable with higher income individuals, while the "skimmed milk" would be those unable to pay the cost of their treatment. Cream skimming can lead to a degradation of public services, for two reasons:

- By distilling out the profitable sections of the market, private providers may interrupt the cross-subsidies that allowed to sustain higher standards of care for the unprofitable sections (see also Chapter 2, section 2.4);

- The 'cream' is usually also the informed part of the market that can put pressure for a general service improvement (in Hirschman's terminology, by exerting 'voice', rather than 'exit'). Once they are gone, the public provider is left with a pool of weaker and less informed clients, which may accept worse service conditions.

Phenomena of cream skimming inevitably occur when the health market is liberalized. Regulation of contracts between purchasers and providers and of the private market for care can correct risk-selection, or cream-skimming resulting in under-treatment of patients. However, such corrections entail tempering individual incentives and restoring central control. Again, in disciplining the competitive rules in the health delivery market, purchasers need to find a balance between the need to maintain uniform or at least minimal standards of care, and that of enhancing results-orientation and performance thorough competition.

Annex 4.1: The Debate on Resource Prioritisation Policies

Beginning in the late '60s, early '70s, a concern for the provision of primary care services, and against an exclusive focus on hospitals and specialized care, progressively emerged in developing countries, as exemplified in the Alma Ata declaration (WHO, 1978)[110]. Then, during the '80s and '90s, a consensus progressively emerged within the health research community to utilize burden of disease and cost-effectiveness as the principal criteria [111] to determine priorities for public health intervention. On the basis of such criteria, the World Development Report (1993) advocated that priority for public funding be assigned to a minimum 'package' of cost-effective services (Basic Benefit Package, or BBP).

The 1993 WDR's recommendations gave rise to several thorny issues concerning the definition of priorities for public financing and provision, which have not yet been resolved. In the debate that followed the publication of the WDR, two among several others points of critique centred on the WDR's lack of attention for:

➢ Implementation/incentives issues. Theoretical cost-effectiveness (MICE) computations used in the WDR were abstract, derived in ideal settings, and ignored a host of complicated feasibility/implementation/incentive issues. Filmer et al. (1998) showed that if one considered the evidence on actual public sector cost-effectiveness (PSCE) from developing countries, one would reach estimates of cost effectiveness several orders of magnitude higher (worse), than those estimated by theoretical MICE computations. Thus, part of the international research effort after the publication of the WDR has tried to articulate and test new ways to improve efficiency and quality of public delivery in developing countries.

➢ The role of the private sector, and the public-private interaction. Policy recommendations in the WDR 1993 failed to recognize the simple reality that health interventions by the public

---

[110] Traditionally, public intervention in health was focused on expanding the delivery system, through physical infrastructure investment and increase in the number of doctors and nurses.

[111] Medical intervention cost-effectiveness (MICE), the principal criterion, ranks interventions according to their ability to achieve a certain outcome (lives saved or disability-adjusted life years gained) per unit cost, according to best practice protocols. Burden of Disease (Body) studies provide estimates of how important different health disorders are (for instance in terms of Daly's lost) to the total BD Criteria for essential package was, in the end, a combination of magnitude of burden + MICE of interventions. However, Gwatkin's work on disease burden in poorest groups showed that these two criteria led to different conclusions.

sector do not take place in a vacuum. The recognition of the private/public differential ability in providing/financing different health services has lead to revisit the criteria for determining priorities.

This angle of looking at priorities for public intervention, sometimes called the welfare economics perspective, bears important consequences for equity. For instance, even if most primary care and outpatient services are in general more cost-effective than most inpatient care services[112], the potential for cost-recovery is much greater for the former than the latter. Catastrophic episodes that require inpatient care are in general the less predictable health events and the more costly to treat. In the absence of private insurance coverage and/or direct subsidization of hospital care, the majority of the population could not afford them.[113]

A recent contribution by Musgrove (1999) tried to reconcile the principles of welfare economics and those of cost-effectiveness. Musgrove provides a decision tree on the rational use of public financing in the health sector based on the two set of principles. [114] His contribution revamped the debate on prioritization criteria, and other scholars have forcefully proposed again the argument that cost-effectiveness should not be a criterion used to determine public resource allocation priorities [115]. Recently, Filmer and Hammer (2002), using the criterion of welfare loss/market failures, plus equity considerations, recommend that priority should be given to (i) pure public goods, (ii) services with large externalities (such as water & sanitation, immunization services), and (iii) services for catastrophic illness.

---

[112] Medical treatments for most catastrophic illnesses tend to occupy low levels in priority scales based on MICE analysis, as they are relatively more expensive and rarely resolving (see Hammer and Berman, 1995).
[113] In the language of welfare economics, the private market equilibrium is characterized by market failures that entail welfare losses.
[114] Mangrove's (1999) decision tree starts with the welfare economics questions, by asking if the service proposed for public funding is a public good, or it bears significant positive externalities. If it does, the next step is to rank it vis-à-vis other services with the same characteristics in terms of cost effectiveness or benefit-cost analysis. In contrast, if the service is question is a private good, or does not entail significant externalities (such as most curative care services), the decision tree asks whether risk of catastrophic costs are involved and whether the proposed beneficiaries are poor or not. Again, all interventions for which the answer is positive are ranked according to cost-effectiveness.
[115] Jack (2000) points out that cost-effectiveness is a criterion to compare alternative services, once it has already been decided that those services are worth of social protection and therefore that the government should subsidize them (because they are non-rival or non-excludible, because there are significant externalities involved, or to correct insurance market failures, or because the services create a pro poor impact, or, more in general, because health is considered as a human right, or a merit good. Merit goods are goods that the government should subsidize beyond the level that individual preferences would dictate. The concept was initially developed by the German school of Public Finance, and later introduced into the mainstream of Public Economics by R. Musgrave.

Chapter 5: Optimal payment systems?

5.1 Introduction

This chapter presents a survey of the theoretical literature that has investigated the main payment (or reimbursement[116]) systems, or PS, for health providers and the incentives associated with each of them. The second part of this Chapter investigates the properties of a specific second-best optimal PS in a context of asymmetric information. A PS is defined as a set of rules defining the criteria according to which a provider of health services receives financial compensation. This can be linked to inputs, the number of patients, their diagnosed health related condition, the type and the cost of care given, a combination of the above, and/or other factors. PS determine the way financial resources are distributed among provider facilities and individual health care providers.

PS play a particularly important role in the health sector, due to the peculiar interplay of demand and supply. Probably more than in any other market, when demanding health services individuals are often largely uncertain about the appropriateness of the service they are receiving. Any health service is an experience good, a good whose appropriateness and quality can be ascertained only ex-post. Even ex-post, the appropriateness of any individual health service is difficult to assess precisely because the patient often receives a multidimensional service (diagnosis, advice, treatment...), and because the final outcome, 'good health' or `health improvement`, which is what people really demand, depends on factors other than the kind of care received, the output produced. The lack of information and the particular psychological condition of patients produce crucial consequences for the interaction between demand and supply. It is the doctor, and not the patient, who generally chooses the amount and the type of services that the latter receives. [117]

---

[116] As we will not specifically be concerned in this work with the issue of who actually pays for health services, whether the government, a social insurance fund, or competing private insurers, we will not explicitly distinguish between payment and reimbursement.

[117] We can better understand how peculiar is the interplay between demand and supply in the health sector by contrasting it with the ordinary interaction that takes place in other markets. Consider for example what would happen if at the onset of each winter consumers had to buy new cloths, naked and just knowing that they feel cold, but without being aware of the quality and the price of the options available in the market...and what makes demand even weaker in the case of health is that winters are not generally foreseeable.

Hence, market outcomes in health are largely determined by the incentives, as well as the cultural influences and professional ethics on the basis of which providers act, without strong influence from the demand side.[118] Payment and reimbursement criteria, in turn, are critical to determining those incentives, and that is why different payment systems can lead to completely diverse outcomes, in terms of accessibility, quality and cost of services offered within the health sector.

5.2 Main PS

In this section, we list the principal payment systems for providers and introduce some of the incentives associated with each of them, as illustrated in the economics literature. First, it is important to distinguish between PS for individual providers and those for provider facilities (such as hospitals, clinics or health centres). Some PS are the same for both institution facilities and individual providers, and, in general, the criteria for paying the latter are very much influenced by those for the former. Yet, this is not necessarily always the case. For example, in several countries of Western Europe physicians employed in hospitals and other facilities continued to receive seniority-based salaries even after the reimbursement criteria for their provider facility had been radically changed, from an input towards an output-based PS.

5.2.1 Main PS for health provider facilities

Focusing first on facilities, until two decades ago across OECD and developing countries one could find two prevailing PS:

- In public health systems provider facilities were mainly reimbursed according to inputs and historical expenditure. Financing was distributed separately for the different line items of the provider's budget (so much for salaries, so much for capital, so much for consumables, etc.), and the total budget was generally rolled over from year to year with minor adjustments. Because at any level of activity the amount received by any facility depended on its "size" in terms of personnel, capital and other inputs, the incentive was to expand these inputs, beyond the technically efficient level.

---

[118] Similar, although less severe, problems arising from the asymmetry of information between supply and

- In parallel, in the private sector providers were mainly paid retrospectively (that is, with full cost-reimbursement), on a fee-for-service basis, directly by patients and/or by indemnity insurers. In a fee-for-service PS the revenue facilities (and individual doctors) receive depends on the volume and the value of the services they provide. Thus, those OECD countries, such as the US, where this kind of PS was more widespread, over time experienced a more pronounced growth in health expenditure, in absolute terms as well as a proportion of GDP. One stream in the US health economics literature also underlined that, when reimbursement to providers is activity-and-cost-based (as in a fee-for-service system), there is no incentive to focus on technological progress that could lead to less costly treatments. Providers can gain by making use of ever more costly treatments and equipment, and even by inducing demand and supplying services above the level that would be clinically justified. This phenomenon, named 'supply-induced-demand', was first studied by Evans (1974). Later, Fuchs (1978) emphasized the welfare-reducing shift in the demand curve induced by providers through advertisement, and Robinson and Luft (1985), Dranove (1988), and Mooney and Ryan (1993) gave a theoretical explanation of supply-induced demand, by pointing at the imperfections in the agency relationship between physicians and patients. Over the last three decades the phenomenon of supply-induced demand has also been extensively studied in the empirical literature. The first study was that by Roemer, 1961, who found a strong correlation between supply and utilization of health services (by examining the cross-sectional correlation between bed-supply and utilization, Roemer found that "a bed built is –roughly- a bed half-filled"). Subsequent literature found a positive association between physicians' ownership of testing facilities and the number of referrals for diagnostic tests (see for example, Crane,1992, and Mitchell and Scott,1992), and strong evidence in favour of the supply-induced demand hypothesis. However, much of the literature documenting phenomena of supply-induced demand has recently been questioned for its econometric shortcomings. By using TSLS regression analysis[119], Dranove and Wehner (1994) found evidence in favour of supply-inducement of child-births[120] (!). Thus, despite a good economic understanding of the rationale behind the supply-induced demand phenomenon under fee-for-service PS, its empirical relevance is still controversial.

---

demand occur in other sectors, such as the financial investment or legal advice markets.

[119] Instrumental variables are used to take into account the endogeneity problem of the explanatory variables.

[120] This lead them to question the identification of the demand inducement econometric models, as they found that the "exogenous" instrumental variables used in the first stage of the TSLS estimation, which are the same as those traditionally used in most empirical works aimed at estimating supply-induced-demand, in fact are not exogenous.

In the last twenty years, in all OECD countries and some developing countries new and more sophisticated PS have been introduced, as part of RAP reforms. The general trend has been to move away from input-based or fee-for-service PS, towards case-based PS, such as DRG-based payments, or cost-and-volume and block payments (particularly for government-owned hospitals).

At this point, a brief explanation of the terminology used in the literature on PS is needed. Output-based PS can be fee-for-service, or case-based, or cost-and-volume based. Case-based (cost-per-case or fixed-price, according to US terminology) PS, first introduced in the US Medicare system (1983) and later on adopted in several other countries, link payment to hospitals to the number and the severity of the cases they treat. Each patient is classified in a specific "diagnostic" group according to his/her principal diagnosis and, correspondingly, a fixed reimbursement is given to the hospital for treating the patient. When this PS was introduced, all pathologies were classified in 470 different Diagnostic Related Groups (DRG); since 1983 the number of DRG categories has been expanded several times.[121] In the late '80s and '90s, case-based PS were adopted in several developing and former socialist economies (FSE) to pay for inpatient care (the first government to adopt them in Latin America was Brazil and among FSE was Hungary; see Chapter 7), in some cases with the same classification system (DRG) as in the USA, in other cases with other classification systems (for example, the "Nosology-based" system in several of the Former Soviet Union countries). Lower income countries (one example is Poland) have sought to experiment with case-based PS based on simpler classification than the DRGs, to reduce administrative costs, prevent manipulations, and provide stronger incentives for cost-control.

With a cost-and-volume PS, similarly to a cost-per-case payment, providers receive additional funding for each extra case they treat, but additional units of service are reimbursed at a progressively lower unit price, according to a schedule agreed upon in advance (see second part of the chapter).

Finally, a fixed-budget, or lump-sum, or block payment gives the health provider facility a fixed budget for the provision of a pre-specified volume of services over a certain period of time, determined ex-ante and completely independent at the margin of the number and the actual cost of the services provided.

---

[121] Presently, DRG codes are based on the 9th revision of the International Classification of diseases.

Fee-for-service and input-based payments are frequently referred to as retrospective PS, because the cost of services is completely reimbursed ex-post. On the opposite side of the spectrum, lump-sum or "block" payments are referred to as purely prospective, because the amount given is fixed ex-ante. At times, cost-per-case or fixed price payment systems are also referred to as prospective payments. In fact, these PS are only partially prospective, because the level of funding per "unit of service" is fixed, but not the total revenue that the provider unit receives, which is dependent on its volume of activity.

Moving from fee-for-service, towards case-based, cost-and volume payment systems and block payments progressively increases the part of providers' payment that is fixed ex-ante, thus increasing also the financial risk associated with unexpected increases in the volume or/and the cost of services (see analysis, hereafter).

5.2.2 Main Payment Systems for individual health providers

The PS for provider facilities described in the previous section can also be used to pay individual providers. For example, the payment correspondent to the input-based PS for facilities is that based on salary for individual providers. In this case, the input paid for is time. Until two decades ago, this was the prevailing PS for individual providers throughout the world in government delivery systems. Health workers received a fixed salary mainly dependent on their seniority, and they were included in the civil service. By contrast, in countries where private provision and financing prevailed, fee-for service PS were mainly used also for individual providers. In recent years, a PS for individual providers and specifically primary care doctors which has become more prevalent is the capitation PS. This consists in a lump sum payment for each patient enrolled with a particular physician, which can be adjusted in relation to age, sex and past medical record of each patient. Two variants are used to determine enrolment: in the first, each patient resident within a specific geographical area is automatically enrolled with the same physician; in the second, each patient can choose his/her primary care physician, also among those practicing outside his/her area of residence.

5.3 The Theory of Regulation and Its Relevance for the Analysis of Reimbursement Systems in Health

The impact of the change from retrospective to prospective PS (refer to section 4.4.1 for an explanation of the meaning of retrospective and prospective PS) have been explored at length

in economic studies on other sectors, such as water, gas, electricity and telecommunication, within the so-called new theory of regulation (Laffont and Tirole, 1993). Ma wrote in 1994: "The change from retrospective reimbursement to prospective payments in the health sector resembles the change from rate of return regulation to price-cap regulation in the telecommunications industry and in other public utilities" (Ma, 1994, p. 96). Hence, before directly examining the change in providers' incentives in the health care sector, it is worth briefly mentioning the main results of the new theory of regulation concerning PS.

5.3.1 Price-Cap (PC) Regulation (Littlechild, 1983)

Under PC regulation, the regulator sets a maximum for the prices of a particular basket of goods that the regulated firm supplies. The regulated firm is able to supply any quantity of those goods at a price equal or lower than the PC. The major advantage of price-cap regulation is that providers can fully internalise the savings from any cost reduction they achieve, thus creating incentives to enhance productive efficiency. In the '80s and '90s PC regulation has been adopted in the UK and in several other countries to regulate the public utility sectors.

One particular way to set the PC is that suggested by Shleifer, in his model of Yardstick Regulation (1985). In synthesis, Shleifer argues that by imposing a fixed price for the product of any industry, based on the **average costs of production** in the industry, the regulator can reproduce conditions of perfect competition, where each firm in the industry faces an infinitely elastic demand at the prevailing market price, and minimizes costs taking those prices as given. Thus, yardstick regulation would achieve productive efficiency and first best equilibria.

To function according to Shleifer's theoretical model, the industry or sector subject to yardstick regulation needs to satisfy the following prerequisites:

- The number of firms active in the industry must be large enough such that no one is able to influence the yardstick price by colluding with the other firms.
- Shocks idiosyncratic to each firm's cost must not be significant. Thus, the mix of inputs and the production processes used by the different firms in the industry must be similar.
- The product must be homogeneous or, at least, consumers can easily ascertain the

quality of the differentiated products/services.

In addition, the regulator must commit not to rescue firms in difficulty, and to let them go bankrupt in case of insolvency.

Some of the above conditions are normally violated in practice. Specifically, the main issues brought out by the literature that has investigated "real life" examples of PC regulation, mainly in the public utilities' sectors, concern the following aspects: a) how to set the price-cap, so that it is at the "appropriate level" and does not recreate the same distortions in production as the Rate of Return regulation did (mainly overcapitalization, as the allowed rate of return was dependent on the capital investments); b) how to weight the different items that are part of the regulated basket of goods or services; and c) how, and how frequently to change the PC over time so that it adequately reflects changes occurring in the industry.

a) To set a PC, ideally the regulator should have detailed information on the industry costs (average costs in case of Schleifer's model). However, this cost information is not available. Thus, in fixing the PC, the regulator needs to make an assessment, based on an imperfect knowledge of some of the regulated companies' cost structure obtained through specific surveys, or based on reported costs by the regulated firms, or on some subjective assessment. If the cap is set too high (relative to the first best level), firms are able to gain extra-profits. In extreme situations, the price-cap may de facto become ineffective, in the sense that prices prevailing in the absence of any regulation would in any case have been lower. By contrast, if the price-cap is set too low, production may become unprofitable, and the PC prevents new investors to get into the market or induces existing ones to leave.

b) Because of administrative and information constraints, each price cap generally groups under the same basket different products from production processes, which are related but still distinct. The ceiling on prices is generally calculated as a weighted average of all the prices of these different products. The theory indicates that neutral weights, which are weights that the regulated firms cannot influence -such as the share in total revenues of each of the regulated products in the basket, ought to be used. However, revenue information is in practice not used, again because of lack of market information and other regulatory constraints. Then, weights are often based on the proportion of allocated inputs (such as labour) to produce each item in the basket. These inputs, though, are a strategic decision variable for the regulated firms, and so distortions in the production process are likely to arise

after the PC is imposed, to loosen regulatory constraints. Such distortions may be more severe if firms also produce products not subject to PC regulation, and whose productions process is closely related to those who are.

c) A separate set of dynamic issues arises in the context of a Price Cap regulation. The PC needs to be adjusted over time, to reflect inflation and changed conditions in the industry. According to the literature, the PC should be revised using a price index (PI), calculated as follows: PI < RPI - X, where RPI is the Retail Price Index, and X is a parameter fixed by the regulator to reflect cost-saving innovations in the industry.

Note that the length of the period between two consecutive adjustments is a critical factor: if it is too long, firms may over-exploit favourable changes in the industry environment, such as a technological revolution, or may be unduly penalised in case of unfavourable shocks; if it is too short, the regulated firms do not have any incentive to reduce costs or innovate, because this would result in a tighter price regulation (**ratchet effect**). When short-term rules are utilized to discipline a long-term relationship, another possible adverse effects is the "**hold up**" **effect** (Hart and Holmstrom, 1987). This is the regulated firms' tendency to uphold investments which could reduce recurrent costs (such as investments in new machineries). Regulated firms tend to under-invest if the value of the investment cannot be recaptured, once it is "sunk", when they fear that at the next round of negotiations their PC will be reduced, because of the recurrent cost savings obtained through the capital investment.

5.3.2 Are the conclusions of the PC and yardstick regulation literature applicable to health markets?

Several contributions in the recent health economics literature (reviewed in the following sections) have proposed the use of case-based PS for health services and goods on the basis of theoretical conclusions similar to those reached by Shleifer (1985) and the theory on PC regulation. The fixed-price PS would allow providers to fully internalize any cost reduction/increased efficiency, and recreate perfectly competitive conditions in the regulated health delivery market. However, as we shall see, the same constraints that prevent PC or yardstick regulation to achieve first best results in other sectors are relevant in the health care sector. Some of the most severe constraints are the following:

- Costs (marginal and average) in correspondence to different levels of production are extremely difficult to compute for health care services. PCs are calculated on the

basis of in depth studies of a limited number of providers (as in case on the DRG system introduced in Italy in 1994), or on the basis of International prices (for example in determining price caps for pharmaceuticals in West Europe).

- Shleifer's results are critically based on the hypothesis that the product is homogeneous or that quality is easily verifiable. This should be contrasted with the importance of the quality dimension, and at the same time the uncertainty that surrounds its measurement in the health care sector. If quality is not verifiable, or it is not enforceable[122], introducing Price Cap regulation can provoke extremely negative results. Providers compete by lowering their quality standards, instead of eliminating inefficiencies, as the model of yardstick competition predicts.[123]

5.4 Payment Systems: Major Trade-offs

The following are the main trade-offs associated with providers' payment systems:

- between creating the incentive to enhance quality, and to contain costs;
- between the objective of avoiding risk-selection and that of stimulating productive efficiency.
- between the aim of containing providers' surplus, and that of inducing them to supply an optimal level of services.
- between administrative accessibility and convenience, and ability to flexibly respond to individual circumstances and provide more complete information.

As we shall see, the "optimal" payment system finds a second-best equilibrium –or a balance- between the above trade-offs. In the rest of this chapter, we will analyze in detail how the theoretical literature has dealt with some of the above trade-offs. Let us just mention here that other dilemmas - not analysed here- may also be relevant. For example, for PS in health care there are dynamic issues similar to those mentioned in section 5.3.1, point c above, related to the optimal length of the regulatory "contract" that disciplines the PC.

---

[122] This is the case when quality clauses of health contracts would be observable by both purchasers and providers, but are not verifiable by third parties and therefore not enforceable by courts.

[123] Moreover, in presence of asymmetry of information on providers' efficiency, fixed-price reimbursements may induce providers to retain higher than optimal surpluses. See model hereafter.

5.4.1 Quality Enhancement - Cost Consciousness Trade-off

In the RAP reforms implemented, in order to create the proper incentives for cost-control both purchasers and providers are increasingly subject to prospective funding, and they have to bear the extra-costs associated with unexpected increases in the volumes of activity and/or in the unit cost of treatment. Unfortunately, there is no single reimbursement system that is able to provide incentives for cost containment, without at the same time creating the incentives for quality skimping.

Several models in the literature have analysed the problem of finding the PS able to "optimally" balance the provider's (the agent) cost reducing and quality enhancing efforts, in the context of a "hidden action" or "(supply side) moral hazard" principal agent model. These models include the following: Ellis and McGuire (1986); Ellis and McGuire (1990); Allen and Gentler (1991); Glazer and McGuire (1994); Hodgkin and McGuire (1994); Ma (1994); Rogerson (1994); Chalkley and Malcomson (1995 and 1998); and Ma and McGuire (1997).

We can graphically represent the position where the different PS stand along the lines of this trade-off:

Figure 5.1: Providers and Type of Reimbursement
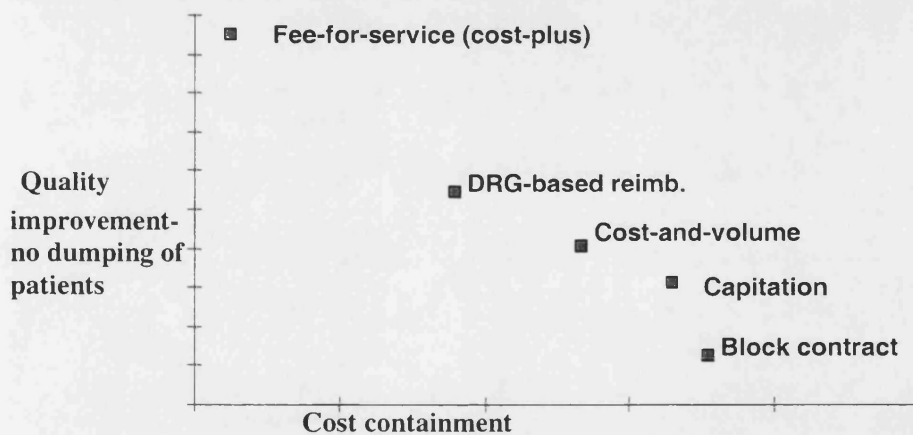
As the above figure shows, moving from fee-for-service towards DRG-based reimbursement, cost-and volume contracts, capitation funding and block contracts progressively increase the part of providers' payment that is fixed ex-ante. The incentive to contain costs increases, but so does the incentive to skimp on quality, and dump patients who are exceedingly costly to treat.

139

To understand the basic forces at play, we will refer to the simple formalisation developed by Hodgkin and McGuire (1994). Their model represents a health provider's quality maximization problem, assuming that its objective function directly considers both quality (intensity of treatment, I) as well as profits (or financial surplus, denoted by the letter "B"). Profit is equal to revenue minus cost, plus any exogenous payment Y (such as donations) the provider receives. Revenue depends on the volume of services produced, X, as well as on their unit price, p (assuming this price is paid by a third-party public or private insurer or purchaser). In turn, the unit price p consists of a lump-sum transfer, α, and a partial cost reimbursement, βc, with $0 \leq \beta \leq 1$. Following Ellis and McGuire (1986), the expression "$1-\beta$" is named "supply-side cost sharing", the fraction of unit costs directly borne by the provider. Finally, both the volume of services supplied and their unit costs positively depend on quality (intensity of treatment). Thus, similarly to the model in Chapter 3, in this model it is assumed that the provider always needs to increase costs in order to improve quality.[124] Following are the basic equations of Hodgkin and McGuire's (1994) model.

$U = U(B, I)$ $U_B, U_I > 0; U_{BB}, U_{II} < 0$ (4.1)

$B = R - C + Y$ (4.2)

$R = pX$ with $X = X(I)$ $dX/dI = X'(\ ) > 0$ $X'' < 0$ (4.3)

$p = \alpha + \beta c$ $0 \leq \beta \leq 1$ (4.4)

$C = cX$ with $c = c(I)$ $dc/dI = c'(\ ) > 0$ $c'' > 0$ (4.5)

The provider chooses quality (I) in order to maximize its objective function:

Max $_I U\{B, I\}$ = max $U_I\{[(\alpha + \beta c(I)) - c(I)]X(I) + Y; I\}$

The equilibrium condition for an optimum is that:

$(dB/dI) U_B + U_I = 0$

That is:

$[(p-c)dX/dI - X(dc/dI)(1-\beta)]U_B + U_I = 0$

Hodgkin and McGuire identify two different effects of any payment system:

♦ $(p-c)dX/dI$ = VOLUME EFFECT. By this channel, the provider is induced to increase the quality of services supplied any time the payment it receives for each case treated is higher than its marginal cost.

- $X(dc/dI)(1-\beta)$ (SUPPLY-SIDE) MORAL HAZARD EFFECT. By this channel, the provider is induced to reduce the quality of services supplied any time the payment it receives for each case treated does not fully cover the cost of providing it.

The supply-side moral hazard effect implies that, as providers bear an increasing share of the (marginal) cost of providing quality, they tend to skimp on it. As pointed out by Ellis and McGuire (1986), moving towards more prospective payments increases the degree of 'supply-side cost sharing', and thus can lead to reductions in equilibrium quality of services.

Note that according to Hodgkin's and McGuire's model the profit or not-for-profit nature of providers is important because it influences their objective function (increasing or decreasing their genuine concern for quality, captured by the term U_I). A benevolent hospital will do exactly what a welfare maximizer agent would want it to do with a fixed budget (this being the definition of "benevolence" in this context). In this case a pure block contract achieves first best results (Chalkley and Malcomson, 1998).

Finally, an important contribution in the literature on optimal payment is that by Chalkley and Malcomson (1995, 1998), on the basis of Ellis and McGuire (1990 and 1993), and Ma (1994). In the context of a multi-task, "hidden action" principal-agent problem (Holmstrom and Milgrom, 1992), they analyse the problem of providing incentives for cost containment and quality enhancement. They prove that, by relying on the 'volume effect', a cost-per-case or fixed-price payment system can induce providers to choose both an efficient (first best) cost-reducing effort, as well as an optimal (first best) quality of services. [125] The "volume effect" (or "demand channel", as Chalkley and Malcomson (1995) rename it, pointing that its magnitude depends on the sensitivity of demand to quality, dX/dI) can induce health providers to maintain services at an optimal quality level. [126] However, the "demand channel" does not work under the following circumstances:

- if there are capacity constraints on the supply side;
- if patients' perception of quality is uncertain;

---

[125] In fact, the above models apply to contracts in health a well-established result in agency theory, which states that when the agent is risk neutral, in hidden action models it is optimal to pay the agent so that he/she/it fully internalises the principal's objective function, and then to redistribute any surplus through lump-sum transfers (Loeb and Magat, 1978).

[126] We precisely show the effects of the "volume effect" or "demand channel" in Annex 5.1.

- if there is a moral hazard effect on the demand side (see Annex 2.2. Patients tend to overconsume health services when insured because they do not face the full marginal cost) that makes it optimal to explicitly ration the volume of services;
- if patients are constrained in their choice of providers; or
- if the volume of demand for treatment and quality are substitutes in patients' demand (for example, if patients make up for a lower quality by consuming more, for example by visiting doctors more frequently; see Ma and McGuire, 1997).

In normal circumstances one of the above constraints applies, and so the only way to achieve an optimal level of quality of services is to directly reimburse part of the quality enhancing costs sustained by providers. Thus, the optimal payment system is in general not fully prospective or fully retrospective. We summarize some of the key results of the theoretical literature on providers' optimal payment systems in the table that follows:

Table 5.1: Optimal PS under different hypotheses on market structure and information

	Provider	Demand	Technology and cost	Optimal payment system
1	Benevolent and risk-neutral	Known	Known	Block contract
2	Benevolent and risk-neutral, but limited capacity to borrow to cover a financial shortfall	Unknown	Known	Block contract + cost-per-case up to the number of patients that the purchaser wants to be treated
3	Self-interested and risk neutral	Responds to quality in a desirable way and is able to pay for quality improvements	Known, with no capacity constraint, and depending on a hidden action effort variable	Cost per case (Fixed price)
4	Self-interested and risk-neutral	Does not respond to quality either because of information asymmetry or because no other option is available		Mixed payment, with a prospective as well as a retrospective component in order to encourage both quality enhancements and cost reduction
5	Self-interested and risk neutral	It is optimal to explicitly ration treatments to contain demand's moral hazard	Capacity constraint is binding	Mixed payment, which partially reimburses Mixed-payment expected rise in cost or
6	Self interested and risk-neutral		Capacity constraint is binding	Mixed payment
7	Risk averse			Mixed payment
8	Risk neutral		Cost is not observable and depends on variable that providers cannot influence, at least in the short run	Non-linear mixed-payments, in order to induce providers' to reveal their cost and to reduce information rents.

5.5 The optimal PS when costs are not observable

5.5.1 Introduction

Building upon Baron and Myerson (1982) and Laffont and Tirole (1993), in this section we analyse the characteristics of an optimal PS when there is asymmetric information in cost variables that are not under the direct influence of providers. These variables may depend on the average severity of illness of the pool of patients that seeks care from each provider, or on the quality of training for staff and administrators, or of the infrastructure). We show that under these circumstances the optimal payment is not linear in the number of cases treated (it is not fixed-price), and must be dependent on the provider's reported cost.

The aim of the model we propose is to contribute to the literature on optimal PS or "contracts" in the health sector (see Malcomson and Chalkley, 2000). In this literature, the term contract is used loosely, to indicate an agreement (for example between a purchaser of health services and a provider), where payment from the purchaser is made contingent upon the services the provider will supply over a certain time period.[127] The distinction between a real contract (such as that between a private insurer and a provider), and a service or regulatory "agreement" imposed by a government purchaser to a health care provider is blurred, and the term contract is used to indicate both situations. Using the more rigorous terminology on PS introduced in the first part of this Chapter (see section 5.2.1), the model that we present investigates the optimal properties of an output (or activity) based PS offered to a private provider, when the purchaser is a public agency at least partially interested in pursuing social goals.

5.5.2 The Model

Consider a situation in which a health care purchaser makes a "contract" proposal to a hospital. This proposal defines a payment schedule t(x), function of the level of services (or number of patients) "x" that the hospital managers agree to deliver (treat). Suppose x is a routine surgical operation. The purchaser's proposal is a "take-it or leave it" offer.[128]

---

[127] See also discussion on purchasing and contracts in Chapter 4.
[128] The hypothesis is that the contract cannot be renegotiated ex-post.

Assume that ex-ante, when the contract proposal is devised, the purchaser does not know the hospital's level of efficiency, or the local conditions that may influence provision costs. For example, costs may depend on ex-ante capital investments that cannot be varied over the contract period and that cannot be verified by a third party in case of dispute, or may depend on patients' average severity of illness when they are admitted to the hospital, which is known only to hospital managers.[129]

Assumption 1: Formally, let us assume that the hospital is characterized by the following cost function [130]:

$$C = C(x, \beta) = \beta + (1 + \beta)x \tag{5.1}$$

where β is a cost parameter known only to the hospital managers (we shall refer to it as the hospital "type"), which influences both fixed and variable costs, and x is the level of services ($x \geq 1$).

Assumption 2: The provider managers know the value of β, but the purchaser managers only know that β is uniformly distributed on [0, 1].[131]

5.5.2.1 The Objective Functions of the Hospital and the Purchaser

The health economics literature that has dealt with agency problems usually models the purchaser as either a profit or a welfare maximizer agent, depending on its ownership status, and the provider as maximizing a single objective function, generally consisting of a combination of profits and other

---

[129] Or, alternatively, costs may be adversely influenced by the strength of local Labour Unions, or by other local input markets conditions that the provider knows but cannot influence, and that no third party is able precisely to verify. Chalkley and Malcomson (1999) and De Fraja (2000) also consider situations where cost of treatment is influenced by severity of illness.

[130] Note that $C_{\beta x} > 0$. This property, known in the literature as "single-crossing property", states that a hospital with a lower β is able to increase the supply of services at a lower marginal cost than a hospital with a higher β. It implies that as the number of treatments supplied increases, cost differentials among hospitals of different type widen. This assumption allows avoiding "pooling" solutions, in which hospitals with different characteristics are treated equally.

[131] Note that the uniform distribution satisfies the monotone hazard rate property: $F(\beta)/f(\beta)$ is non-decreasing in β. $F(\beta) = \beta$, $f(\beta) = 1$, $d[F(\beta)/f(\beta)]/d\beta = 1$.

variables that are positively correlated with patients' welfare (such as quantity and quality of services) [132]. For simplicity, we will assume that:

Assumption 3: the provider's objective function coincides with that of its managers, and the latter maximize profit or financial surplus. [133] If the hospital receives a payment ("transfer") equal to t(x), and it incurs a financial cost equal to $C(x,\beta)$, managers' utility is the following:

$$U_H = s(t,x,\beta) = t(x) - C(x,\beta) = t(x) - \beta - (1+\beta)x \qquad (5.2),$$

Assumption 4: The purchaser's objective function is increasing in the level of services (number of patients) that the provider delivers (treats), and it is decreasing in the payment given to the hospital:

$$U_P = V(x) - t(x) = K\log(x)^{[134]} - t(x) \qquad \text{with } K > 1 \qquad (5.3),$$

The above characterization of the purchaser's objective function can be justified for several reasons, for example: 1) because the purchaser maximizes consumer surplus, which in turn could be positively associated with the level of services; or 2) because of positive waiting lists for the routine surgical treatment x, and pressure from the central regulatory agency or the central government to contain them. [135]

---

[132] There is an ample literature that has specifically studied hospitals' objective function. It can be roughly divided in two strands. The first one has modelled the hospital as a unitary decision entity, whose behaviour can be reduced to the maximisation of a single objective function, whilst the second one has considered explicitly the different agents present within the hospital, whose interests are partially in conflict with each other. In the first strand, we find Rice (1966) (non-profit hospital maximising quantity), Newhouse (1970) and Lee (1971) (doctors maximising quality). In the second one, the most important contribution is perhaps that by Harris (1977), who criticises the unitary representation of hospitals' behaviour and explicitly represents the conflict between specialists and managers. Specialists behave as agents of patients, and their objective is to maximise intensity of treatment. Managers, on the contrary, are interested in maintaining the hospital financial equilibrium, which depends on the payment the hospital receives. Thus, there is generally a conflict between the resources that specialists demand and the amount that managers are willing to provide.

[133] In case of public hospitals, it would be realistic to assume profit maximisation as long as this would allow managers to pursue other objectives, such as the enlargement of the hospital, increase of managerial perks, or peaceful relations with the hospital personnel.

[134] Note also that gross welfare associated with treating x patients is concave in x. To motivate the concavity assumption, consider that, given a certain forecast of the population's needs for x, the purchaser's management plans to first satisfy the more urgent cases. Then, as x increases, the marginal utility of additional services diminishes. Also, the purchaser's managers might be increasingly concerned with budgetary limits. Note that we do not assume, as it is normally done in the literature, that the purchaser considers the shadow cost of collecting public funds.

[135] Or the purchaser could also want to expand production beyond current demand. Consider for example the situation in many poor countries or rural areas, where people may not be informed about the potential benefits of any particular health treatment. In this case, the public agency which purchases services may desire a level

5.5.3 Results:

a) The Optimal Contract When Cost Is Observable

Let us first consider the case when the value of the cost parameter β is observable. The purchaser's preferred value of x, denoted by x^{FB}, is defined by solving the maximization problem:

$$x^{FB} = \text{Arg max }_{x,t}[K\log(x) - t(x)] \qquad (5.4)$$

Subject to the constraint that the hospital must achieve at least a minimum financial surplus (Participation Constraint or PC):

$$U_H = t(x) - \beta - (1+\beta)x \geq 0, \qquad \text{for each } \beta \text{ and } x \qquad (5.5)$$

When cost is observable the PC (5.5) holds with equality, because the purchaser is able to reduce payment to a level just sufficient to cover the cost of service provision. Then, the maximization problem becomes:

$$\max._x E\{U_p\} = [K\log(x) - (\beta + x(1 + \beta)) - U_H] \text{ with } U_H = 0 \qquad (5.6)$$

The first-order-conditions define the following FB level of activity[136], as a function of β:

$$x^{FB}(\beta) = \frac{K}{1+\beta} \qquad \text{for each } \beta \in [0, 1], \qquad (5.7)$$

Proposition 1: If $\beta' < \beta''$, $x^{FB}(\beta') \geq x^{FB}(\beta'')$. When cost is known, in equilibrium a less efficient hospital (higher β) treats fewer cases than a more efficient hospital (lower β).

Proof: $dx^{FB}(\beta)/d\beta < 0$

Proposition 2 indicates that the above optimal contract with perfect information can be implemented by simply paying specialists according to a linear function of the number of patients that demand treatment (according to a fixed-price PS).

---

of services superior to existing demand.

[136] We will assume throughout the chapter that the value of K is large enough so that gross marginal utility is larger than marginal cost for values of $x \leq 1$, that the necessary first order conditions for a local maximum are always sufficient conditions for a global maximum, and that they are satisfied for positive values of x. The

Proposition 2: The FB level of activity $(x=x^{FB})$ can be implemented by a contract in which payment to the hospital $t(x)$ consists of two parts:

- a fixed lump sum transfer;
- a fixed-price (or cost-per-case) payment per patient treated:

$$t_x = V_x (x^{FB}) = \frac{K}{x^{FB}}$$
\hfill (5,8)

The proof is straightforward (see Annex 5.1), and it relies on the 'volume effect' described in section 5.4.1. This result was first proved by Chalkley and Malcomson, 1995.

b) The Optimal Contract When Cost is Not Observable, and Cannot be Varied Over the Contract Period

In case of incomplete information, the purchaser faces an adverse selection problem, deriving from its ignorance of the hospital cost parameter, the need to preserve the hospital budget balance, and hospital managers' interest in exploiting the information asymmetry.

In this situation (when the purchaser's managers do not know β), it is not efficient to link payment to the cost reported by the hospital managers, unless the purchaser devises a payment schedule that makes it always convenient for the latter to report truthfully. In fact, following Baron and Myerson (1982), to find the second-best contract, one can apply the revelation principle and represent the interaction between the purchaser and the hospital as a direct revelation Bayesian game, where the strategy of the latter is a mapping from the set of possible cost-types into itself [b(β), b being the cost parameter that the hospital managers report to the purchaser managers, and β the true cost parameter]. The purchaser's strategy (named 'revelation mechanism') is a collection of proposals (t(b), x(b)) specifying a payment and a certain number of cases, functions of the cost the hospital managers report. The revelation principle guarantees that the equilibria of the direct revelation game coincide with the equilibria of the original game, if we focus on 'incentive compatible' revelation mechanisms, where the hospital managers truthfully report the cost parameter β. The incentive compatibility or truth-telling (IC) constraint can be written as follows:

---

equilibrium volume will be the first integer number below the equilibrium x.

$s(\beta, \beta)^{137} = t(\beta) - C(x(\beta),\beta) \geq s(b,\beta) = t(b) - C(x(b),\beta)$, for each$\beta$,   (5.9)

The timing of the revelation game is as follows:

<u>Time 0</u>: Nature draws a cost parameter from $F(\beta)$ and reveals it to the hospital managers.

<u>Time 1</u>: The purchaser devises a second best (SB) contract (that defines a certain number of cases and a corresponding payment, both functions of the cost parameter β that hospital managers report). Following a standard result in the literature, we can solve the purchaser's maximization problem as an optimal control problem (all the proofs are confined to Annex 5.2).

The equilibrium value of $x = x^{SB}$ is defined by the first order conditions, valid for each β:

$$V_x(x^{SB}) - C_x(x^{SB}, \beta) - (\frac{F(\beta)}{f(\beta)})C_{\beta x}(x^{SB}, \beta) = \frac{K}{x^{SB}} - (1 + \beta) - \beta = 0 \qquad (5.10)$$

, which give the following second best level of activity, function of β:

$$x^{SB}(\beta) = \frac{K}{1 + 2\beta} \qquad \text{, for each } \beta \in [0, 1], \qquad (5.11)$$

Proposition 4. Under asymmetric information, the optimal level of activity for each provider type is a non-linear decreasing function of the cost-type β reported by the latter, according to equation (5.11).

Let us analyse the shape of the optimal level of x as a function of β. In equilibrium the purchaser's marginal expected benefit from purchasing an additional unit of service from a provider of type $[\beta,\beta+d\beta]$ (the probability that the provider's cost parameter is indeed between $[\beta,\beta+d\beta]$ is $f(\beta)d\beta$), must balance the additional expected payment to the hospital. Such marginal payment consists of two parts: first, the expected payment to compensate the hospital for the cost of providing an additional unit of service. The second is an incentive to offset the hospital managers' interest in misreporting costs. This second component gives rise to what is known in the language of regulation theory as an 'information rent' or surplus. Such surplus increases as x increases, because the cost

---

[137] The first parameter is the cost parameter revealed by hospital managers, while the second is the hospital's true cost-parameter. Assuming that the incentive compatibility constraint is satisfied, from now on we will denote $s(\beta,\beta)$ simply by $s(\beta)$.

differential among hospitals of different type widens ($C_{\beta x} > 0$). Thus, to limit the information rent, the purchaser penalizes those hospital managers who report relative higher cost by reducing the level of services to be purchased from their hospital. As a result, in equilibrium all hospitals but those whose managers declare being of the most efficient type are kept at an activity level lower than the first best level ($x^{SB}(\beta) < x^{FB}$ for each β except for $\beta = 0$).[138]

Proposition 5:

a. In correspondence to the optimal contract, all hospital managers except those who report their hospital being the least efficient, enjoy a positive net surplus. Such surplus (named "information rent") is equal to:

$$s^{SB}(\beta) = \int_\beta^1 \frac{dC}{d\beta}(x^*(\beta), \tilde{\beta})d\tilde{\beta} = \int_\beta^1 (1 + \frac{1}{1+2\beta})d\tilde{\beta} = 1 - \beta + \frac{1}{2}\log 3 - \frac{1}{2}\log(1+2\beta)$$

(5.12)

b. The total payment that the hospital receives is equal to: $t^{SB}(\beta) = C(x^{SB}(\beta),\beta) + s^{SB}(\beta)$.

Note that the optimal x^{SB} is a monotone decreasing function of β. Thus, it is possible to invert it, and to obtain a function $\beta=\beta(x)$, which is also monotone decreasing, and write the optimal payment as a direct function of x.

$$T(x^{SB}) = t(\beta(x^{SB})) = C(x^{SB}, \beta(x^{SB})) + s^{SB}(\beta(x^{SB})) = \frac{x+3}{2} + \frac{1}{2}\log 3 + \frac{1}{2}\log x$$

(5.13)

Equation 5.13 shows the characteristics of the optimal payment system under conditions of asymmetric information. It shows that higher activity levels come with higher total budgets (so the optimal payment is dependent on level of service). By deriving T(x) with respect to x one obtains:

$$\frac{dT(x^{SB})}{dx} = \frac{\partial C}{\partial x} + \frac{\partial C}{\partial \beta}\frac{\partial \beta}{\partial x} + \frac{ds^{SB}}{d\beta}\frac{\partial \beta}{\partial x} = \frac{x+1}{2x}$$

(5.14)

---

[138] By doing so, the purchaser makes it more costly for efficient types to mimic less efficient service providers.

Proposition 6. With asymmetric information over cost, the optimal payment schedule is non-linear in x: $dT(x^{SB})/dx$ is decreasing in the volumes purchased.

With asymmetric information over costs, a cost-per-case or fixed-price payment system (which is linear in activity levels) would not be optimal because it would not respect the incentive compatibility constraint, and would thus induce any hospital manager to report a cost level higher than the true one.

Equation (5.14) implies that additional levels of service (or number of patients) are compensated at decreasing unit rates. Mathematically, T(x) is a concave function of x:

$$\frac{dT^2(x^{SB})}{dx^2} = \frac{-1}{2x^2} < 0 \qquad\qquad (5.15)$$

Note that in correspondence to any incentive-compatible mechanism, being $ds^{SB}/d\beta = -\partial C/\partial\beta$ (see equation 5.18 in Annex 5.2), the last two terms of equation (5.14) cancel out. Then, $dT(x^{SB})/dx$ is equal to $\partial C/\partial x^{SB}$. In words, expression (5.14) indicates that in correspondence to the optimal contract, each provider receives as marginal reimbursement a sum corresponding to the marginal cost of the "marginal" provider-type at each level of activity. The "marginal provider-type" is the provider that is assigned –and self-selects- precisely that specific equilibrium level of activity (and corresponding payment) along the optimal payment schedule.

5.6 Conclusions

When devising new payment systems for health providers, health reformers may have indulged in the idea that all valuable objectives (productive and allocative efficiency, effectiveness, quality of services, and equity) could be achieved simultaneously with a simple PS, if only the right system could be found. However, the existing theoretical literature tells us a rather different story. It points at the fact that in most situations a choice has to be made among different objectives, in the sense that those PS that are more adequate to achieve some of them are completely incapable, or even harmful, to achieve others. We have also shown that the optimal PS and the impact of any PS on quality and other relevant variable depend on market conditions, on the information that agents can have access to, ad on the variables they can influence. For example, the "demand channel" (patients'

positive response to quality improvements) may be used to induce providers to offer higher quality services, but in a context characterized by capacity constraints this "demand channel" cannot work. If the theoretical analysis is correct, then one must draw the conclusion that a PS that is optimal in any circumstance in fact does not exist, but it is really dependent on the priority objective(s) one chooses. Our analysis also indicates that the accompanying institutional and informational framework is crucial in determining the performance and the constraints associated with any PS. Payment systems ought to be used flexibly, to be properly matched to the changing priorities and market conditions, and be accompanied by effective supervision to minimize the distortions they may cause and to fully bring out their potential.

In the second part of the chapter we described the characteristics of the optimal PS, or, by using the terminology of the theory of regulation, "contract", between a purchaser and a provider of health services when costs are not observable. The model builds upon the analytical framework first proposed by Baron and Myerson (1982) and Laffont and Tirole (1993). The second best contract identifies a level of services (number of patients) to be treated and the corresponding payment as functions of the cost of provision reported by the hospital managers, and it displays the following features: 1) the volume of services purchased from the hospital is inversely related with the latter's cost; 2) in general, it is lower than the full information level of services; 3) the hospital managers enjoy a positive information rent in equilibrium, which increases with the level of services that the hospital supplies; 4) more efficient hospitals are assigned a higher total payment than less efficient ones and 5) additional levels of service (number of patients) are reimbursed at decreasing rates. The model we presented is static, and thus it is unable to describe the dynamic interaction that takes place in reality between purchasers and providers, where learning over costs takes place over time and where additional pieces of information enter in subsequent negotiations once they become available.[139] Moreover, the model does not explicitly consider the impact of competition between different providers, which could provide the purchaser with useful information to be used in contracts.[140]

---

[139] In other words, the PS analysed in section 5.5 is optimal, given the prior information of the purchaser, but it is not ex-post efficient, once the cost parameter is common knowledge. The features of the model presented above may be realistic in contexts characterized by rapidly varying costs for exogenous reasons, for instance following a technological shock that rapidly changes the way certain diseases are treated. Or, if overhead costs are significant and providers can easily shift such joint costs from one service to the other, and contracts are dealt separately for each single specialty and service.

[140] In fact, the above model can easily be extended to consider the case in which the purchaser faces many different providers of unknown efficiency. In that case, an important result of the existing literature is that, as long as different providers are subject to the same exogenous shocks, relative performance evaluation can improve the terms of the contract in favour of the purchaser. In fact, the possibility of making comparisons

Despite the above limitations, it is arguable that the need to overcome the existing asymmetry of information concerning costs and other key contractual variables (for instance quality) should become a central concern in shaping real-world contracts between health purchasers and providers. According to the model we described, purchasers should vary the levels of services demanded from each provider in order to elicit information about their cost level, and providers that are (relatively) less efficient should be assigned "contracts" with lower levels of services and lower budgets. The optimal payment system is a cost-and-volume PS, and not fixed-price.

In conclusion, the model presented here has mainly a prescriptive content, but it is nonetheless of interest to see whether some of its "predictions" have been borne out in real-world situations. The evidence from the 1991 quasi-market reforms in the UK (for the UK, see for example, Propper, 1996 and 1997 and Audit Commission, 1997), suggests that: [141]

- after the purchaser-provider split is introduced, purchasers become more cost-sensitive, build up managerial expertise and develop new Management and Information Systems (MIS) that would allow them to compare costs, activities and quality of care of different providers;

- in spite of the strict regulation often existing on the range of reimbursement prices, providers start to use prices strategically, and sometimes accept 'volume discounts' to secure a higher 'share of the market' in exchange for lower (unit) prices;

- as a result, successful providers start achieving surpluses, sometimes well-above the mandatory rates set by regulation;

All these results are predicted by the model that we proposed in this chapter, and suggest that the latter could be considered as a good starting point to further study the purchaser-provider interaction.

---

among different providers would also shift the balance in favour of prospective payments, because the possibility of misstating costs is limited when the purchaser can make comparisons. The model presented could only be used to study a situation in which the number of providers is taken as exogenous. In that case, where health contracts were auctioned among several bidder-providers characterised by different values of β, it would be possible to show that a contract proposal characterised as in the text would be able to select the most efficient provider (McAfee and McMillan, 1986 and 1987; Laffont and Tirole, 1987). A slightly different and more complex formalisation would be required to compare the advantages of introducing many regulated providers, relative to a bilateral monopoly situation (two recent articles in which market structure is endogenised are Auriol and Laffont, 1993, and McGuire and Riordan, 1995).

[141] Refer to Chapter 6 for a throughout review of the evidence available from other countries on the impact of the purchaser-provider split, and the new PS.

Annex 5.1 Proof of Proposition 2

To induce the provider to choose to treat the purchaser's preferred number of cases, it is necessary to structure the payment system in such a way that the provider maximizes its surplus by choosing $x = x^{FB}$. x^{FB} is defined by the following first order condition:

$$V_x(x^{FB}) - C_x(x^{FB}, \beta) = 0, \qquad \text{for each } \beta \qquad (5.16)$$

In our model, the provider chooses the number of cases to be treated by maximising:

$$\text{Max}_x U_H = s(t,x, \beta) = t(x) - C(x,\beta)$$

Assume also that in correspondence to x^{FB}, the fixed component of payment is adjusted such that the provider's participation constraint is satisfied. When payment is determined as in Proposition 2, the provider's objective function becomes:

$$s(x, \beta) = (\frac{K}{x^{FB}})x - \beta - (1+\beta)x \qquad (5,17),$$

The first order conditions are satisfied by $x = x^{FB}$, the purchaser's preferred number of patients treated. QED

Annex 5.2 Proof of Propositions 4 and 5

To define the optimisation problem we must first redefine the Participation and the Incentive Compatibility constraints of the direct revelation game.

First, consider the Incentive, or truth-telling constraint (IC, equation (5.9) in the text).

We denote:
$s(b,\beta) \mid_{b=\beta} = s(\beta)$; $t(b) \mid_{b=\beta} = t(\beta)$; $x(b) \mid_{b=\beta} = x(\beta)$; $\partial s(b)/\partial b \mid_{b=\beta} = \partial s/\partial \beta$ and $\partial x(b)/\partial b \mid_{b=\beta} = \partial x/\partial \beta$.

From: $s(\beta) = t(\beta) - C(x(\beta),\beta) \geq s(b,\beta) = t(b) - C(x(b),\beta)$, $\forall \beta$, by differentiating $s(b,\beta) = t(b) - C(x(b),\beta)$ at the truth-telling optimum (when b, the cost parameter reported equals the true parameter β), and using the envelope theorem, we derive the following two conditions for the IC to be satisfied:

1) A necessary condition for the incentive compatibility constraint to hold:

$\dfrac{\partial [t(b) - C(x(b), \beta)]}{\partial b}\Big|b = \beta \equiv 0$, from which one can derive the Necessary Condition:

$$\frac{ds}{d\beta} = -\frac{\partial C}{\partial \beta}(x(\beta),\beta) = -(1 + x^{SB}(\beta)) \qquad (5.18),$$

2) The first order condition stated above must hold for all possible βs. We can therefore differentiate it with respect to β, obtaining (the second term is derived applying Young's theorem):

$d(F.O.C)/d\beta = [S.O.C.] - [C_{\beta x}(dx/db)_{b=\beta}] = 0.$

To satisfy the second order conditions, the first term must be negative. The second term must then be positive. Given that $- C_{\beta x}$ is negative (hypothesis 2a), $dx/d\beta$ must also be negative. Thus, a sufficient condition for the IC constraint to be respected is the following:

$$\frac{\partial x}{\partial \beta} \leq 0 \qquad (5.19)$$

Second, consider the Participation Constraint (PC) (5.5). Given that $C_\beta > 0$, $s(b,\beta)$ is monotone decreasing in β. Thus, if the PC is satisfied for $\beta=\beta^1=1$, it is also satisfied for all other β, lower than β^1. Moreover, since leaving payments to the hospital is costly, the PC constraint will be satisfied as an equality at $\beta^1=1$:$s(\beta^1=1) = t(1) - (1+2x) = 0$.

Hence, taking into account the IC and the PC constraints, the purchaser's maximisation problem can be rewritten as follows:

$$\max_{x(\beta)} \int_0^1 (\log x(\beta)) - [1 + x(\beta)(1+\beta)] - \int_0^1 [1 + x(\beta)] \, d\tilde{\beta} \,)2d\tilde{\beta}$$ (5.20)

subject to:

a) $U(\beta^1) = t(\beta^1) - C(x(\beta^1), \beta^1) = t(1) - (1+2x) = 0$

$$b) \left. \frac{\partial U}{\partial \beta} \right|_{b=\beta} = -\frac{\partial C}{\partial \beta}(x(\beta), \beta) = -(1 + x^{SB}(\beta))$$

$$c) \frac{\partial x}{\partial \beta} \leq 0$$

Let us temporarily ignore constraint c). We will explain the general approach to this maximization problem, and then just state the results with our particular specification of the objective and cost functions (equation 5.10 in the text). From:

$$s(\beta) = s(\beta^1) - \int_\beta^{\beta^1} \frac{ds}{d\beta} d\tilde{\beta}$$ (5.21),

where β^1 is the upper limit of the cost parameter (and β^0 is the corresponding lower limit), considering the PC and the necessary condition for the IC constraint, one can write:

$$s(\beta) = \int_\beta^{\beta^1} \frac{dC}{d\beta}(x(\beta), \beta) d\tilde{\beta}$$ (5.22)

Thus, the objective function of the purchaser becomes:

$$\text{Max}_x = \int_{\beta_0}^{\beta^1} [V(x(\beta)) - C(x(\beta), \beta) - \int_\beta^{\beta^1} \frac{dC}{d\beta}(x(\beta), \beta) d\tilde{\beta}] dF\tilde{\beta}$$ (5.23)

There are two equivalent methods to solve problem (5.23):

Method 1:

We can integrate by parts the last term of equation (5.23), thus obtaining the following expression:

$$\int_{\beta_0}^{\beta^1}\int_{\beta}^{\beta^1}\frac{dC}{d\beta}(x(\beta),\beta)d\tilde{\beta}]dF(\tilde{\beta}) = \int_{\beta_0}^{\beta^1} F(\beta)[\frac{\partial C}{\partial \beta}(x(\beta)\beta)]d\tilde{\beta} \qquad (5.24)$$

Then, $\int_{\beta_0}^{\beta^1}\int_{\beta}^{\beta^1}\frac{dC}{d\beta}(x(\beta),\beta)d\tilde{\beta}]dF(\tilde{\beta}) = \int_{\beta_0}^{\beta^1} F(\beta)[\frac{\partial C}{\partial \beta}(x(\beta)\beta)]d\tilde{\beta}$, the objective function of equation (5.23) becomes:

$$\int_{\beta_0}^{\beta^1}\left[V(x(\beta)) - C(x(\beta),\beta) - \frac{F(\beta)}{f(\beta)}\frac{\partial C(x(\beta)\beta)}{\partial \beta} \right]dF(\tilde{\beta}) \qquad (5.25)$$

Finally, by inserting the values of V(.) and C(.) in the text (respectively, expressions 5.1 and 5.4), we obtain the first order conditions in equation (5.10) and (5.11).

Method 2:

We can apply optimal control theory to solve maximisation problem (5.23) subject to constraints a), b) and c) in equation (5.20). The Hamiltonian is the following:

$$H = [V(x(\beta)) - C(x(\beta),\beta) - s(\beta)]f(\beta) - \mu\frac{\partial C(x(\beta),\beta)}{\partial \beta} \qquad (5.26)$$

where $\mu(\beta)$ is the Pontryagin multiplier.

Taking 'x' as the control variable, and 's' as the state variable, we obtain by the maximum principle:

$$(5.27)$$

$$\frac{\partial \mu}{\partial \beta} = -\frac{\partial H}{\partial s} = f(\beta)$$

The boundary of s(β) at $\beta=\beta_0$, the lower bound, is unconstrained. Hence the Trasversality Condition at $\beta=\beta_0$ ($\mu(\beta_0)s(\beta_0) = 0$) implies: $\mu(\beta_0) = 0$, and integrating (5.27), one obtains: $\mu(\beta) = F(\beta)$.

Finally, by substituting V() and C() with the values in the text (expressions 5.1 and 5.4), and by maximising the Hamiltonian with respect to x(β), we derive equations (5.10) and (5.11).

Let $x^{SB}(\beta)$ be the Second Best solution. The last constraint we have to check in maximisation problem (5.20) is constraint c), i.e., $dx^{SB}/d\beta$ must be non-increasing in the hospital's type. The implicit function theorem implies that:

sign $dx^{SB}/d\beta$ = sign $d(F.O.C._{x})/d\beta$

$$sign \frac{dx}{d\beta} = sign - C_{x\beta} - (\frac{F(\beta)}{f(\beta)})C_{x\beta\beta} - C_{x\beta} \frac{d \frac{F(\beta)}{f(\beta)}}{d\beta} \tag{5.28}$$

Given $C_{\beta x}>0$, the above expression is always negative if $F(\beta)/f(\beta)$ increases as β increases, which is true in case of the Uniform distribution.

Chapter 6: The Equity Impact of Resource Allocation and Purchasing (RAP) Reforms

6.1 Introduction

This chapter studies the equity impact of resource allocation, strategic purchasing and payment system (RAP) reforms in health. In the first part, the chapter reviews the evidence on the distribution of health and health care utilization across socio-economic groups in developing countries. Such evidence shows a striking consistency in the association between poverty and poor health, and that the distribution of health benefits across socio-economic groups is highly unequal, favouring the wealthier segments of the population. Then, the chapter investigates into the meaning of equity in health: we argue that the concept of an equitable distribution of health resources is not as straightforward as it first appears. Hence, assessing changes at the margin may be difficult, and different equity principles and measures would lead to opposite conclusions.

In the second part, the chapter presents the empirical literature that has investigated the impact of RAP reforms from an equity perspective. We argue that some RAP reform components, such as the new resource allocation mechanisms, produced a strong pro-poor impact, but that the evidence of the impact of other components, such as the new contracting mechanisms and payment systems for providers, is much more mixed, and that these may well have generated adverse equity impact.

One of the main conclusions of the paper is that the evidence available on the RAP reforms' impact on equity is largely inconclusive. The lack of a scientifically rigorous evidence base to evaluate the RAP reforms is partially due to intrinsic measurement difficulties. Aggregate measures of equity, especially those concerning distribution of health outcomes, change slowly, while most RAP reforms in developing countries are quite recent. However, the lack of conclusive evidence is also caused by an insufficient attention to monitoring and evaluation the equity dimension of the RAP reforms. For instance, one can hardly find any data to evaluate the impact of RAP reforms where the change in utilization of essential health services is measured disaggregated by socio-economic groups. We hope that in the future new and more conclusive evidence will soon be brought to bear on some of the tentative results presented here.

6.2 Evidence on Existing Inequalities in Health

The evidence available on distribution of health and health resources across socio-economic groups is presented in studies for the most part not related to the evaluation of specific health reforms. We briefly present this evidence in this section.

Regardless of the diversity of the existing studies in terms of measurement approaches to equity and poverty, study design and geographical focus, the empirical evidence indicates a striking consistency in the association between poverty and poor health. Beaglehole and Bonita (1997, p.1) assert that poverty is the most important cause of preventable death, disease and disability.

Inter country comparisons present evidence on health inequalities across countries, and their evolution over time. Some of the key findings are the following:

- The bulk of mortality and morbidity affects people living in the developing world. 98% of the deaths between birth and 15 years and 83% of the deaths between 15 and 59 years occur in the developing world (Murray and Lopez, 1996).

- These inequalities have not decreased over time. Gwatkin, 2000b, analyzing World Bank data set, concludes that in the period 1970-1990 infant mortality differentials across countries declined in absolute terms, but that they increased in relative terms ("While infant mortality in 1970 was around 6.5 times as high in the poorest as in the richest countries, it was over 11 times as high by 1990 (Gwatkin, 2000b, p. 19)". He also points at the fact that, unlike under-5 mortality figures, adult mortality differentials did unequivocally narrow over time.

- Over 30% (53% in sub-Saharan Africa) of deaths in the developing world occur in children younger than 5 years. In terms of DALYs lost the three main groups of disease are perinatal disorders, lower respiratory infections, and diarrhoeal diseases, all of which are prevalent in the developing world, especially in the poorest countries and among the younger and poorer segments of the population.

- Life expectancy is strongly associated with income per capita. However, the relationship is not linear (Preston, 1980) and is characterised by high variability, as different countries with similar level of income achieve widely different results in mortality rates and life expectancies (see Chapter 4, Box 4.1).

More recent literature has looked into the relationship between poverty and health within countries. The main findings of this literature are three: A) mortality is related as much to socio-economic inequalities (in asset ownership, expenditure, income, education attainment, etc.) within countries, as it is to differences in absolute wealth/income amongst them; B) within each country, mortality rates are positively correlated with levels of deprivation (for example, Anand and Ravallion, 1993, show that a significant part of the impact of income per capita on mortality withers away once we explicitly consider the share of the population in poverty as an explanatory variable); C) there is increasing evidence that demonstrates an independent positive impact of social capital on health outcomes (see the evidence presented by Subramanian et. al., 2001; Kawachi et. al., 1999).

Investigations of health differentials across socio-economic groups in low and middle income countries are based on two major sources of evidence: Living Standard Measurement Surveys (LSMS) and Demographic and Health Surveys (DHS). While LSMS contain very accurate information of households' level of consumption and expenditure, there is less detailed information on health seeking behaviour and health outcomes. By contrast, DHS data contain very accurate information over health status and health service utilization, especially for reproductive, maternal and child care services, but no information on households' income or expenditure. The two above main sources of evidence are discussed respectively by Gwatkin et al. (2000) and by Wagstaff (2000).

On the basis of DHS data collected in over 40 developing countries, Gwatkin et al. (2000) analyze inequalities across socio-economic groups in: a) infant and under 5 mortality; b) levels of malnutrition; c) incidence of diarrhoea and acute respiratory infection; d) fertility rates; e) information on HIV/AIDS; and f) use of Maternal and Child Health services. They adopt a wealth index developed by Filmer and Pritchett (1998)[142] to divide the population into different wealth quintiles according to reported asset ownership, and show that disparities between poor and non-poor vary enormously across countries. The tables below present some of their key findings. On average across countries, a child born in a household belonging to the lowest wealth quintile is roughly twice as likely to die before reaching 5 years, than a child born in a household from the highest wealth quintile. Inequalities are highest in the Latin American/Caribbean region.

---

[142] Filmer and Pritchett (1998) use information on household wealth and assets from DHS questionnaire and develop a wealth index using through principal components analysis. They then apply this index to the study of education attainment in 35 countries.

Table 6.1 and 6. 2: Intra-Country Disparities in Infant and Under-Five Mortality[143]

6.1. Infant Mortality

Region	No. of Countries	Poor-Rich Ratio		Concentration Index[144]	
		Mean	Range	Mean	Range
Sub-Saharan Africa	21	1.67	1.11 to 2.46	-.081	-.003 to -.141
Asia/Near East/North Africa	9	2.33	1.42 to 3.93	-.125	-.051 to -.195
Latin America/Caribbean	11	2.66	1.26 to 4.18	-.145	-.043 to -.251
Total	40	1.87	1.11 to 4.18	-.106	-.003 to -.251

6.2 Under-Five Mortality

Region	No. of Countries	Poor-Rich Ratio		Concentration Index	
		Mean	Range	Mean	Range
Sub-Saharan Africa	21	1.79	1.27 to 2.60	-.095	-.040 to -.164
Asia/Near East/North Africa	9	2.69	1.69 to 4.60	-.147	-.084 to -.210
Latin America/Caribbean	11	2.99	1.55 to 4.67	-.167	-.071 to -.259
Total	40	2.06	1.27 to 4.67	-.124	-.040 to -.259

Source: Gwatkin, 2000b

An interesting finding of Gwatkin et al. (2000) is that countries with lower mortality and morbidity rates among children are in general also characterized by wider disparities across socio-economic groups. This finding seems at odds with the non-linear relationship between income and health,

---

[143] Two patterns of the data worth noting are that under-five mortality rates are more concentrated than infant mortality rates (this could result from problems of underreporting of infant mortality), and that the standard deviations are extremely large (so that parameters are not statistically significant for several countries).

[144] The concentration index is equal to: $1 - 2\int_{0}^{1} MR_i(w_i)dw$, where MR is the cumulative proportion of mortality rates among children graphed against the cumulative proportion of their households' wealth (i=1,…,5). A negative (positive) value of the concentration index indicates inequality favouring the rich (poor).

characterized by diminishing returns to scale, and could be explained by one or a combination of the following reasons:

a) Rates of diffusion of health-promoting innovations are different across different segments of society, first benefiting the wealthier segments of society.

b) Increase in per capita income is associated with increases in income disparities.

c) Societies experience varying degrees of health disparity because some of them put in place policies that make the health-income relationship more elastic while others do not.

d) A potential negative health externality occurs associated with being poor in a richer country.

Using LSMS data, Wagstaff (2000b) compares data on infant and under-5 mortality rates from 9 low income countries. His study finds significant inequalities in infant and under-five mortality rates across quintile-expenditure groups in all countries, although the estimates are not as pronounced as in Gwatkin et al. (2000). For six countries Wagstaff's estimates are directly comparable with Gwatkin's.[145]

---

[145] A third set of measures is presented by WHO (1999). The WHO estimates are not based on household data. They extrapolate the mortality rates at different incomes from information on the average mortality and on the distribution of income. The WHO estimates show stronger inequalities than those derived from household data. According to these estimates, those who live in extreme poverty (less than 1 US dollar adjusted for purchasing power parity) are five times more likely to die before reaching the age of 5 years than those above the poverty line, and two and half times more likely to die between the ages of 15 and 49 years.

Table 6.3: Concentration Indices for Infant and Under-Five Mortality within Six Developing Countries as Measured by Different Studies

Country	Infant Mortality		Under-Five Mortality	
	Country Info. Sheets (Gwatkin et al.)	Wagstaff	Country Info. Sheets (Gwatkin et al.)	Wagstaff
Côte d'Ivoire	-0.107	-0.095	-0.115	-0.096
Ghana	-0.093	+0.018	-0.135	-0.028
Nepal	-0.060	-0.109	-0.096	-0.132
Nicaragua	-0.094	-0.150	-0.124	-0.162
Pakistan	-0.051	0.000	-0.084	-0.017
Vietnam	-0.143	-0.009	-0.159	-0.016
Un-weighted Mean	-0.091	-0.064	-0.119	-0.076

Sources: Wagstaff, 2000b; Gwatkin, 2000b

The evidence also indicates that given all the other determinants of health (see Annex 4.1), even large disproportionate increases in the use of services by the poor would produce relatively smaller reductions in health disparities (for example, see Victora, 1999, on the health impact of a MCH program which carefully targeted the poor in the city of Pelotas, Brazil).

Information on health inequalities and on their determinants is much more developed and detailed in high income countries. In general, the literature has found persistent health inequalities, a strong relationship with income inequalities and levels of deprivation, but weak links with specific health systems' characteristics. For example, a recent study on the causes of health inequalities and their link with socio-economic deprivation and health systems' characteristics in industrialized countries by Van Doorslaer et al. (1997) found that the USA have the highest level of health inequalities, followed by the UK, the Netherlands and then all the other West European Countries and Japan, where the extent of health inequalities across socio-economic groups are the lowest. Inequality is statistically significant in all countries. By regressing health inequalities on: a) total pro-capita health expenditure; b) public share in total health expenditure; c) total GDP per capita; d) Gini coefficient

for income distribution, they find that only the last term, income inequality, produces a statistically significant positive association with health inequality, whilst income per capita, total health expenditure and even the public share seem not to matter. Their finding seems to confirm the importance of the "relative" pathway through which income and income distribution may affect poor health (Subramanian and Belli, 2001).

6.3 The distribution of benefits from health care

The evidence presented by Gwatkin et al. (2000) also indicates that the distribution of utilization of health care services is extremely unequal across socio-economic groups. In spite of the fact that poorer segments of any society suffer disproportionably for death and morbidity, the wealthier segments utilize health services both in the private as well as the public sector disproportionably more than the poor. Even those services, such as interventions against communicable diseases, which address diseases mostly concentrated among the poor, are unevenly utilized more by the rich.

The following figure aggregates data for 43 developing countries on immunization of children[146] for households belonging to different expenditure quintile groups. It shows that immunization services, which in principle would provide disproportionate health benefits to poorer households who are currently suffering from communicable diseases, are in fact utilized more by wealthier households. In Sub-Saharan Africa children belonging to the top quintile households are twice as likely to receive immunization services than those belonging to the poorest households.

---

[146] Data presented in: HNP and the Poor: An Integrated Policy Framework for Improving the Outcomes for the Poor, by Soucat and Wagstaff (2001).

Figure 6.1: Inequalities in Immunization Rates between Wealthier and Poorer Households

Immunization Rates among Poor and Rich
(Source: Gwatkin, Poverty Data Sheets)
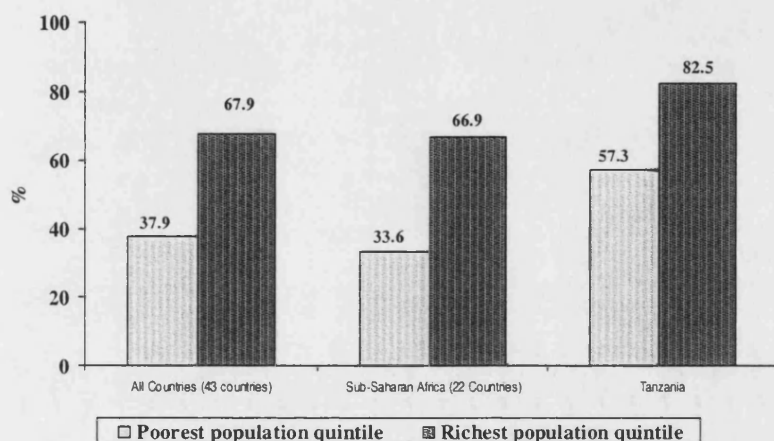
Source: Soucat and Wagstaff, 2001

A series of benefit-incidence studies[147] that focused on the benefits dispensed by public health systems to different socio-economic groups indicated strong pro rich inequalities in the distribution of benefits from government = expenditure on health care services, particularly in Sub-Saharan Africa. A study by Gwatkin (2000b) summarizes the results of this literature as follows:

---

[147] Benefit-incidence studies compute the average subsidy for the different types of care (usually, primary, outpatient specialist care, inpatient care) from budgetary information and from data on service use. They estimate the subsidy going to each group from information on the distribution of utilization by income group of the different service-types.

Mathematically, benefit incidence is estimated by the following formula:

$$X_j \equiv \sum_i U_{ij} \frac{S_i}{U_i} \equiv \sum_i \frac{U_{ij}}{U_i} S_i \equiv \sum_i e_{ij} S_i$$

X_j	=	health sector subsidy enjoyed by group j,
U_{ij}	=	utilization of service i by group j,
U_i	=	utilization of service i by all groups combined,
S_i	=	government <u>net</u> expenditure on service i, and
e_{ij}	=	group j's share of utilization of service i

Thus, incidence studies do not consider therapeutic benefits, nor adjusts for different need across quintiles. They estimate the financial redistribution and do not consider the revenue side. They also do not convey any information about the reasons behind differential use of services across socio-economic groups. For a review of the early literature, see Birdsall and Hecht, 1995

Table 6.4: Percentage of Financial Subsidy from Government Health Services Accruing to Poorest and Richest 20% of the Population: Continent Averages

Region	Primary Care		Hospital care					
			Outpatient		Inpatient		Total Hospital care	
	Poorest Quintile	Richest Quintile	Poorest Quint.	Richest Quint.	Poorest Quint.	Richest Quint.	Poorest Quint.	Richest Quint.
Africa	15 (7)	23 (7)	12 (2)	36 (2)	16 (2)	34 (2)	10 (5)	33 (5)
Asia	21 (2)	16 (2)	7 (1)	41 (1)	5 (1)	41 (1)	13 (1)	22 (1)
E. Europe	16 (2)	22 (2)	--	--	--	--	12 (2)	29 (2)
Latin America	--	--	--	--	--	--	--	--

Total government health care expenditure		
Region	Poorest Quint.	Richest Quint.
Africa	12 (7)	30 (7)
Asia	19 (5)	21 (5)
E. Europe	13 (2)	27 (2)
Latin America	29 (8)	14 (8)

Source: Gwatkin 2000b.

Note: Each figure in parentheses indicates the number of countries included in the continent average that appears immediately to the parentheses' left.

Note that according to the above table only in Latin America government expenditure on health care seems to be pro poor. In Latin America only Brazil shows a pro-rich distribution of public expenditure on government-run facilities However, the estimates available do not take into account the public subsidy to social security, which covers only formal sector employees and that accounts for over half of total public health expenditure in several LA countries.[148] Inequalities in service utilization are starker in the poorest countries of Sub-Saharan Africa. A benefit incidence study in

---

[148] In general in LA public services fulfil mainly the residual role of providing basic services for those unable to pay, whilst the wealthiest segments of society utilize private facilities (see Chile Case Study, Chapter 2). Those employed as civil servants or formally in the private sector are frequently reimbursed by their own employment category's insurance or sickness fund, frequently subsidized by government sources. Moreover, note that because benefit incidence studies do not take into account the revenue side, they tend to overestimate the redistributive impact of government health expenditure. In most developing countries, the main sources of financing it for the government are indirect taxation, payroll contributions and out-of pocket payments, which are all regressive.

seven countries in Africa shows that the richest 20% of the population receive well over twice as much financial benefit than the poorest 20% from all (inpatient as well as outpatient) government expenditure (Castro Leal, F., Dayton J., Demery, L. and K. Mehra. 1997). In all but two of the seven countries (rural Kenya and South Africa), the richest 20% also gain more than the poorest 20% of the population from primary care expenditure. In a study on Indonesia[149], Van de Walle (1998) concludes that the redistributive impact of public health expenditure is negligible. Using benefit incidence measures, only in Malaysia and in the Philippines public health expenditure seem to be pro-poor (Castro-Leal, 2000). In a benefit incidence analysis on India, Mahal et al. (2001) show that the public subsidy for health is regressive, with the rich getting almost three times more than the poor, as the following figure indicates.

Figure 6.2: Share of the Public Subsidy for Curative Care according to Income Group
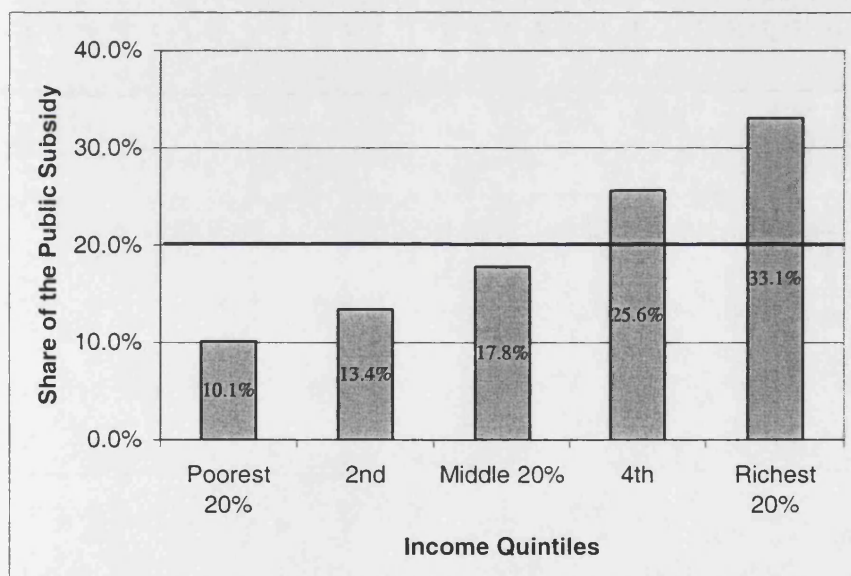
Source: Mahal et al., 2001

Mahal's study in India also indicates that different health services are characterized by an opposite redistributive impact. By measuring the cumulative benefits of different services for different socio-economic groups using a Lorenz curve, Mahal et al. (2001) show that the public subsidy to hospital

---

[149] Note that even if benefit incidence analyses in Asia show less unequal distributions than in Africa, the share of public expenditure over the total is on average much lower in the former than in the latter continent. A larger share funded by private out-of-pocket payments entails a larger burden on the poor, and overall health benefits more skewed in favour of the rich.

care (inpatient and outpatient) is more pro-rich than the subsidy to outpatient care provided by primary health care facilities (PHC) [150], as the following figure shows:

Figure 6.3: Subsidy Benefits Concentration Curve by Type of Care
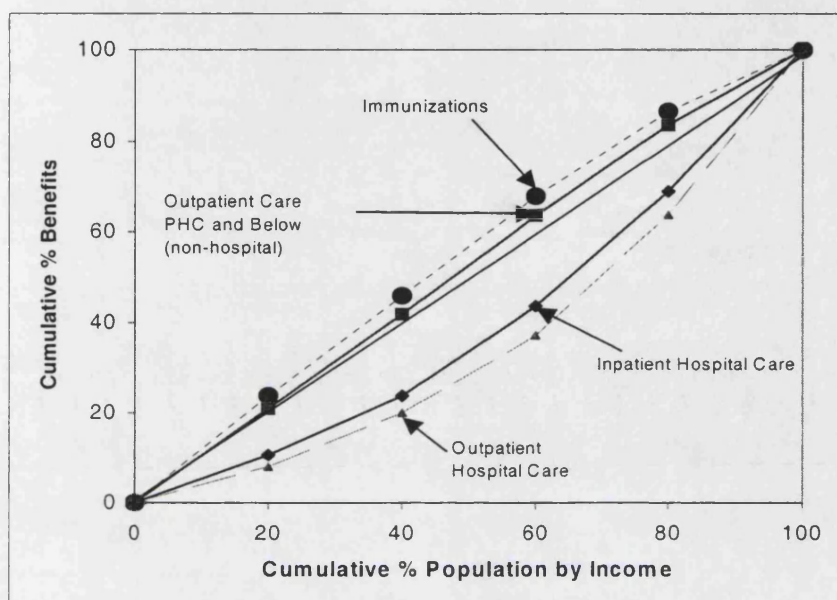
Source: Mahal et al., 2001

The disproportionate use by the wealthier segments of society is due to better information, social norms more conductive to healthy behaviours (such as use of preventive services) and other demand side differences across socio-economic groups (see Annex 4.1). Hence, even if a new health program or service were to be made universally available (so that the poor would not suffer from higher financial barriers to entry, or need to travel longer distances), in general the rich would still end up utilizing the service more than the poor (see, for example, evidence on change in service utilization, after the introduction of a new Maternal and Child health (MCH) program in Ceara, Brazil, presented by Tendler and Freedheim, 1994). However, the higher utilization by the rich is also due to:

---

[150] Within hospital care, inequalities across socio-economic groups are stronger for rural than for urban populations, suggesting that in India the rural poor cannot access hospital care. Also, stark differences exist across Indian states. Kerala is the only state with a pro-poor record for curative services. In three states, Maharashtra, Gujarat and Tamil Nadu, distribution of curative services is not statistically different from equality, whereas in all other states benefits' distribution is strongly pro-rich, particularly in North and North East India (Uttar Pradesh, Bihar).

1) an uneven geographical distribution of public expenditure, health facilities, human resources and other inputs, which are concentrated urban and wealthier areas;

2) high user-fees (for the most part informal or under-the-table (see Chapter 8) for services provided in public facilities.

1. For example, in Bangladeshw and Karnataka, India, the more developed districts receive a higher share of government total expenditure on health, as illustrated in Case Study 6.1.

Case study 6.1: Distribution of health benefits in Bangladesh and in Karnataka, India

A measure of general development is provided by the Human Development Index, which incorporates infant mortality, literacy, and GDP per capita into a single index (1 is the highest level possible, 0 the lowest). UNDP computed HDIs for each district in Bangladesh (UNDP, 1996). The relationship between the HDI and public spending per capita is illustrated in the figure below. The general trend suggests that districts with weaker development, as measured by the HDI, receive lower funding by the Government.

Figure 6.4: Relationship between Human Development Index and Public spending per Capita in Bangladesh (1 BTaka = US$ 0.025-0.030. Bangladesh 1999-2000)
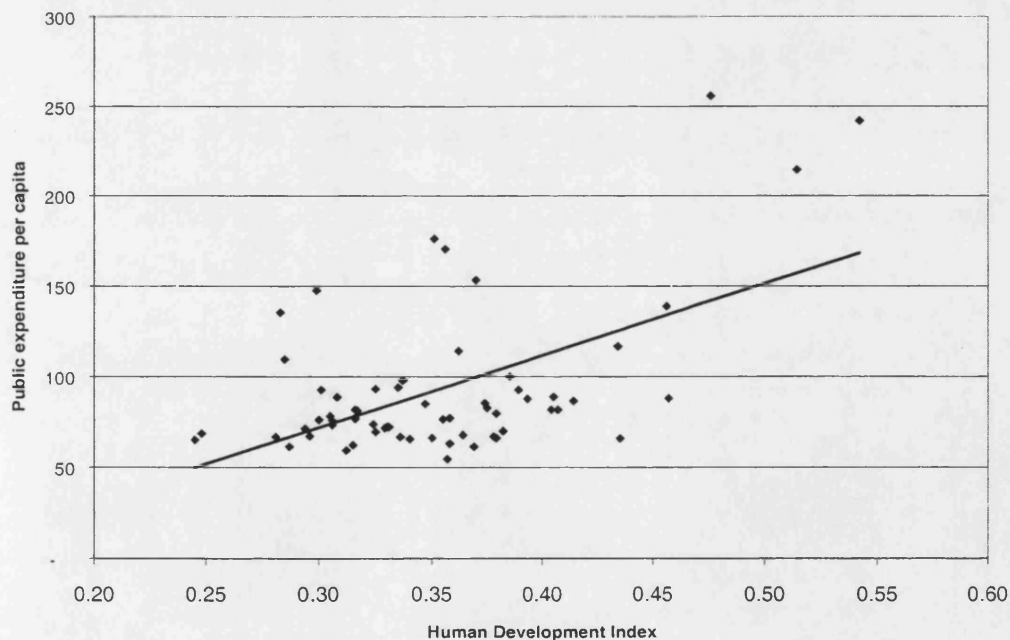
Source: Bangladesh PER, 1999/2000

169

A Public Expenditure Study in the state of Karnataka, India (2004) found similar results. The study indicated that in per capita terms the Government of Karnataka was giving more to relatively more developed districts relative to underdeveloped districts. As Figure 6.5 indicates, the trend marginally improved over time but in 2002 still wide disparities persisted.

Figure 6.5 Per Capita Health Expenditure in constant prices in 9 Districts

	1998-99	1999-00	2000-01	2001-02	2002-03
☐ Total of 2 Dev Dist	69.02	67.24	57.57	66.73	58.35
■ Total 3 Med Dist	44.19	51.99	47.65	51.48	48.99
☐ Total 3 Un dev Dist	35.57	41.22	38.63	44.04	40.08
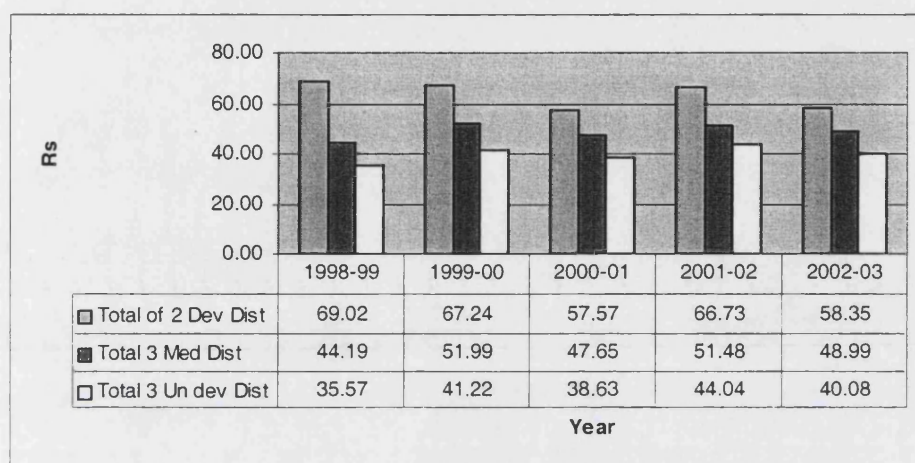
Source: Public Expenditure Review, Health Sector, GoK, 2004

2. Most poor and middle income countries do **not perform well also in terms of financial protection,** particularly against health related catastrophic expenditures. In most poor countries OOP payments constitute the principal source of financing for health care services (see Chapter 8). For example, in India they are estimated to contribute up to 80 percent of the total financial resources for health care. A recent study found that nearly 25 per cent of those hospitalized fell below the poverty line because of huge medical costs (World Bank, 2001). Most of these payments go to private facilities and providers, but there is also evidence of a significant level of OOP payments for services received in government facilities, some of which are unofficial and are not recorded as user-fees revenue. The analysis of the NSS data (1995-96) indicated that nearly 23 per cent of the people interviewed in rural areas cited financial reasons for not availing treatment from government health care unit. This percentage was only 15 per cent in 1986-87. Such high prices discourage a wider utilization of inpatient services.

The inequalities analyzed here for public services in utilization of services across socio-economic groups, because of demand factors, the uneven geographical distribution of facilities and the high OOP payments associated with seeking care (legal user fees or, more frequently, illegal informal payments), are starker in private health care services. These services are almost exclusively financed by OOP payments in low income countries, or private health insurance, in high income countries. Everywhere, higher quality private services, particularly in urban facilities and from formally qualified specialists, are extremely expensive and not accessible for the poor. For example, in Karnataka India, the average price for medical services paid per episode of hospitalization was found to be equal to Rs. 3,291 in a public facility, and Rs.13,313 in a private facility, according to a health utilization and expenditure survey done by the Government of Karnataka. Similar results, with private services costing 4-5 times more than public services, were found in a recent survey in Bangladesh (World Bank, 2005). Thus, even if and where public health expenditure is pro rich, the poor still rely on the public sector more than any other socio-economic group, particularly for essential services, such as deliveries and immunizations. By looking at immunization, attended deliveries and other services, Gwatkin (2000b) shows that the negative outliers in terms of distribution of utilization are countries such as Turkey and Morocco, which mainly rely on private financing and provision. This is illustrated by Mahal et al. (2001) for institutional deliveries in the case of India.

Figure 6.6: Distribution of Institutional Deliveries per 1,000 Births in Public and Private Facilities According to Income Quintile
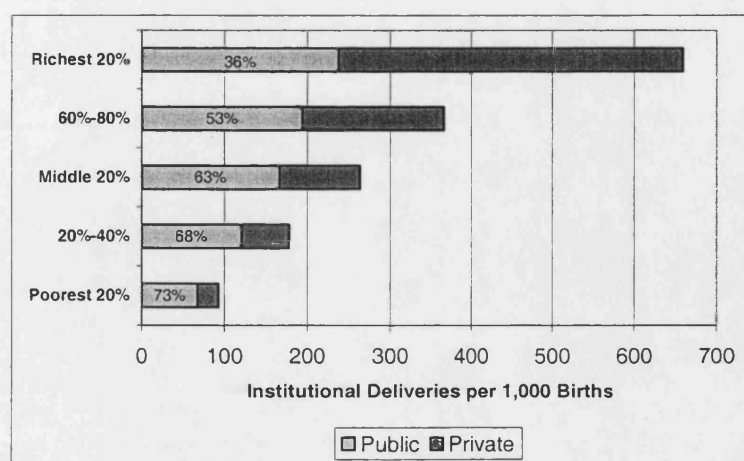
Source: Mahal et al., 2001

Finally, note that the evidence concerning the distribution of health care leads to different conclusions in industrialized countries. These countries achieved a fairly even distribution of the financial burden and of utilization rates across socio-economic groups, although there are still pockets of under-served poor. Van Doorslaer and Wagstaff, 2000 look at the distribution across income groups of utilization and need-adjusted utilization rates across 10 European countries, plus the USA. Their paper highlights the following results:

- Utilization of all services is pro-poor in high income countries. Hospital use, which is pro-poor everywhere, influences this result. However, for some of the countries utilization patters are not significantly pro poor.

- Utilization rates become less pro-poor and even moderately pro-rich once we take into account of increasingly precise indicators of need (need-adjusted utilization rates). Van Doorsaler et al. utilize age, sex, self-assessed health status (SAH) measures and a vector of chronic conditions to adjust for need.

- Utilization is pro-rich for outpatient specialist care. The existence of a GP gate-keeper role does not seem to influence the rate of specialist visits, which remain pro-rich even in countries such as Denmark, Finland and the Netherlands, where patients do not have direct access to public specialist visits.

- The presence or absence of universal public health insurance does not seem to significantly influence the distribution of utilization rates. The US and Switzerland do not emerge as outliers, and in outpatient care show similar pro-rich results as countries such as Sweden and East Germany, characterized by universal insurance.

- For 6 countries, it has been possible to separately analyze GP visits. These turn out to be evenly distributed everywhere, but in Belgium and in Ireland, where they are pro-poor. It is not possible to determine whether the pro-poor distribution in Belgium and Ireland depends on the exemption of poorer individuals from outpatient care co-payments or from other confounding factors.

- During the 1990s inequalities seem to have been widening in some rich countries (for the UK, see: Report of Health Inequalities, 1998). The increasing health inequalities are generally ascribed to the rising income and other socio-economic inequalities, and not to health system's specific variables.

6.4 What Is the Meaning of an Equitable Distribution of Health and Health Benefits?

In general, the literature on equity and health has taken a strong egalitarian position[151], refuting market allocation mechanisms as a benchmark to follow for health services, and advocating a strong role for the government to re-equilibrate the distribution of resources for health care, both across countries and within each country in favour of the poor and those in more need. In the measures more widely utilized in this literature, such as the concentration index, any departure from the state of perfect equality in the distribution of resources for health is considered as inequitable. In assessing how much *additional care over and above* what the individual could him/herself afford or choose, this literature has presented several principles of equity, based either on equality of access, or equality of utilization, or (re)distribution of health care treatment according to need. In the following section we briefly present these principles.

By contrast, a minority position within the literature claimed that the extent of market correction and of individual protection should be limited to a minimum also in the health care sector. In poorer countries this minimum would include essential services such as maternal and child care, perhaps some emergency services, but should exclude most other health treatments (World Bank, WDR 1993). Allocation of resources in the health care sector should be left mainly to individual choices and market allocation criteria.[152]

A more detailed discussion of the assumptions at the basis of these two opposed positions concerning equity in health is presented in Annexes 6.1, 6.2 and 6.3.

---

[151] For a careful and clear review of the existing measures of health inequalities, both in utilization and outcomes, see Wagstaff and Van Doorslaer, 2000.
[152] For example, proponents of the Public Choice Theory (Brennan and Buchanan, 1977) underscore that it is not at all guaranteed that public intervention will achieve more equitable results than markets. According to this theory politicians and civil servants are in general self-interested individuals who utilize their public offices to consolidate their personal power and wealth, and their decisions hardly reflect their constituencies' welfare.

6.5 Equity in health care. The different egalitarian principles

Equality of access and equality of utilization: According to Le Grand (1982), potential equal access to treatment means that individuals should face the same money and time costs for using health services. Mooney (1983) also defends this principle, claiming that it is more respectful of individual preferences than the principle of equality of utilization or treatment: faced with the same costs of accessing treatment, individuals may still choose different levels of treatment. By contrast, Olson and Rodgers (1991) argue that equal access actually means that everybody can potentially consume the same amount of health services. Hence, opportunity costs must be the same (travelling time and other costs included) across income groups, and the poor must be subsidized for their health consumption by the rich until their opportunity cost for treatment is equalized. Pushing Olson and Rogers's (1991) argument to its logic conclusion, Le Grand (1991) argues that equality of access ultimately requires that Choice (or Opportunity) Sets be the same for all individuals.

Distribution of health care according to need. Another commonly advocated equity principle for redistribution of health resources is need.[153] This view advocates equal treatment for equal need, and that those who are in greater need should come first. The literature has proposed different definitions of need. Culyer and Wagstaff (1993) summarise them as:

- need as severity of illness;
- need as ability to benefit;
- need as the minimum amount of resources required to exhaust capacity to benefit.

Each of the above definitions of need leads to different conclusions concerning the equitable distribution of health care services. Assuming for the purpose of illustration Quality Adjusted Life Years (QALYs) as a measure of the quantum of health that each individual did or could enjoy[154], we illustrate some of these conclusions in table 6.5.

---

[153] The literature refers to the two principles of horizontal and vertical equity. The two principles were first articulated by the Greek philosopher Aristotle: the horizontal equity principle states that equals should be treated equally, whereas the vertical equity principle states that those who are unequal should be treated differently. Applied to health provision they are usually translated in the principle that access or use of health services should solely be based on need, and independent of socio-economic condition or of any other non-need related characteristic (such as gender). Applied to health financing they are usually translated in the principle that contribution to funding of health services should be in direct relationship with one's ability to pay.

[154] The principle limitation of QALYs is that they attempt to measure quality of life on the basis of "expert"

Table 6.5: Equity and (re)distribution of health resources

Item	Retrospective health	Prospective health
Health from health care	Previous QALYs gained from health care	Expected QALYs benefit from health care
Health not from health care	Previous QALYs enjoyed without health care	Expected QALY profile without health care

According to the ability to benefit exception of need, in this example we would advocate the criteria on the top-right cell to distribute health services. Equity would demand that expected QALY gains be equalized across individuals. The perspective is forward looking and would give priority to those most able to benefit from health care. There are two exceptions of the notion of "capacity to benefit". One is to consider "marginal capacity to benefit (MCB)". Equalizing MCB across individuals is the principle that maximizes incremental health benefits from health care. This is the most straightforward and widely used efficiency principle in health economics, but it is also sometimes advocated as an equity principle in the literature. In fact, it can lead to extremely inequitable choices, whenever those who are marginally more able to benefit (and should then come first according to this principle) are also the wealthier, healthier, or both.[155]. A second exception is Culyer and Wagstaff's (1993) "minimum amount of resources required to exhaust capacity to benefit". By these means, a consideration of severity of illness is indirectly phased in the definition of need as capacity to benefit, and the results obtained are in general more egalitarian than by just considering marginal capacity to benefit (see Culyer and Wagstaff, 1993).

However, equity criteria based on an exception of 'need' different from 'capacity to benefit' are possible as well, and they are indeed frequently preferred by individuals and households, when asked about priority setting criteria. In Sweden, for example, care for chronically, terminally ill and severely handicapped patients which ranked extremely low in terms of cost-effectiveness criteria were found to be popular and classified as 'essential' care in all household surveys that were conducted to inform the government on people's preferences. [156] Choosing the bottom-right cell,

---

opinion, which can be arbitrary. The welfare (subjective) assessment of benefits, as well the social evaluation of health gains, may be totally different from any "objective" measurement of health gains.

[155] This criterion ominously recalls principles originally proposed by Herbert Spencer. Spencer (1820–1903), a British sociologist, first applied Darwin's theory of evolution to the study of human societies. According to his views, individuals who contribute more to a society and those who are more fit should be preferred over the others.

[156]Swedish Parliamentary Priorities Commission, "Priorities in Health Care: Ethics, economy, implementation", Fritzes, Stockholm, 1995.

where need is interpreted as severity of illness, priority is given to those who would be more severely ill (would enjoy the "least number of QALYs"), without care, regardless of their ability to gain extra health as a result of this care or their 'objective' capacity to benefit.

Another principle would be to use the same criteria accepted for education, where everybody is given a set amount of free services (equality of inputs), and every extra is given according to merit, or it is fully paid for. In the case of health, one should then equalize care received across individuals. Alternatively, one could try to equalize health benefits, or QALYs gained from subsidized care[157] over the life cycle, and choose according to the criterion in the top-left cell, Table 6.5, column 2. Finally, one might want to bring in considerations linked to the age or the past health condition of potential beneficiaries and give priority to the young or to the disabled (those who have enjoyed the least QALYs without health care), according to the criterion in the lower-left cell of Table 6.5, column 2.

Referring back to the discussion of the trade-offs in Chapter 4, note that all of these equity-motivated criteria, with the exception of the marginal capacity to benefit-based redistribution principle, also conflict with any possible effectiveness principle, which would tend to put resources first where the link between inputs and outcomes can be stronger. In other words, there is in general a trade-off between distributing resources for health care to maximize gains in health outcomes, and to achieve greater equity (equity-efficiency trade-off).

All of the above discussion should suggest how complicated any equity assessment is. By adopting any of the above principles-and consequently a different measure of equity or inequity-, one is lead to diverging equity assessments of any given distribution of health or health care resources, or of any given change in such distribution. Case study 6.2 below illustrates a concrete example from Ecuador.

Case study 6.2: The equity impact of social health insurance in Ecuador

Waters (2000) estimates the equity impact of a new social insurance program which covers health services for formal sector employees in Ecuador, and of a plan for its extension to self-employed or to the workers' dependants. He points out that many of the measures currently used to measure "inequity" (see for example, Wagstaff and van Doorslaer, 2000) negatively asses any departure from a state of perfect equality in health status or in need-adjusted utilization of health services, without consideration for the "levels" of care achieved

---

[157] The issue here is not to decide priorities starting from a situation where no services are available; rather, to evaluate how much *additional care* should the state allow an individual to receive *over and above* what he/she could afford to pay for.

by the different groups. Thus, the measures of inequity generally adopted in the literature imply that any pro-rich departure from a state of perfect equality is inequitable, even if the poor are positively affected. By using two of these "egalitarian" measures of equity[158], the concentration index and the Atkinson distributional measure[159], Waters (2000) shows that distribution of utilization of health services in Ecuador was already pro-rich, and it was made even more inequitable by the new social insurance program. The new program would still be an adverse impact on equity even if were extended to dependants or self-employed. The benefits of the social health insurance program would still be regressive, although less so in case the program were extended to the self-employed or to dependants.

However, Waters also considers an alternative view of equity, according to which an intervention that improves coverage for the poorer members of society could be considered as equity-enhancing, even if increased the level of overall variability in health benefits across the population. As a third measure of "inequity" in line with this second view of equity, Waters proposes a weighted welfare function, where the utilization of services by the poorer quintile is weighted more than for the second-poorer quintile, etc. Using this other measure of the degree of inequity in the distribution of health benefits, Waters shows that equity would be unequivocally enhanced by the social insurance program, particularly if the benefits were to be extended to the self-employed or dependants. In conclusion, he shows how different notions of equity in the distribution of health benefits can be translated into different measures, and how these would lead to a different evaluation of the impact on equity of the same interventions.

6.6 Equity. Theory versus Practice

If we compare the concrete evidence on the distribution of health and health benefits in developing countries presented in the first sections with the above conceptual discussion, the contrast between the two is palpable. The conceptual literature on health and equity was developed mostly in Western Europe, and risks of being irrelevant to judge the situation in poorer countries. For example, much of

---

[158] Let us briefly explain a general formulation of the existing measures of inequity, based on the measurement of inter-group or inter-individual differences in access to health. If H_i is a measure of access or health status of socio-economic group i, where i=1,...,I ranks individuals or groups in increasing order according to their socio-economic status (for instance, their income or expenditure), the degree of inequity is measured by a direct function f of the rich-poor differences: Inequity = f [$\sum_i \sum_j (H_j - H_i)$] with j>i. In the need-adjusted measures based on utilization of services, the above formula is replaced by: Inequity. = f {$\sum_i \sum_j [(H_j | N_j) - (H_i | N_i)] j>i$ }, where H_i is a measure of utilization of services and N_i is a measure of health need (for instance, self-reported health status) for socio-economic group i. Paradoxically, using such measures of inequity, a scenario where no person in need could receive any health care would be considered more equitable than a scenario where the poor receive good care, but inferior to that received by the rich.

[159] The Atkinson distributional measure evaluates the total welfare loss from an unequal distribution of utilization of health services.

the above conceptual literature on health and equity is centred around accessibility problems, in a sense giving for granted that health services, once made accessible for all, would also be of acceptable quality standards. By contrast, by looking at the evidence from developing countries we saw that the quality dimension of services is of great importance (see Chapter 4, section 4.3). Government health services may be in theory accessible to (almost) all, but they are of such abysmal quality particularly in the peripheral and poor areas that in reality very few utilize them. In these settings, a focus on equality of treatment, or equal treatment for equal need is more appropriate than just focusing on accessibility.

Second, in analyzing how the current distribution of health and health care benefits can be concretely made more equitable in poor countries, one has to take into account their political economic, institutional and managerial constraints, which are much tighter than in industrialized countries. For example, all the above criteria advocate that priorities ought to be set among individuals and not services, but at best what a government can do in a poorer country is to reprioritize across services rather than individuals, either because of the political or technical impossibility to exclude the non-poor from publicly subsidized services[160], or because of the impossibility of assessing people's socio-economic status or need for health services. Considering these political-economic constraints, some claim that the best strategy to reduce health inequalities would in fact be to expand the scope of universal programs. However, the "mapping" from services to individuals is far from perfect. As section 6.3 pointed out, without any explicit targeting of the poor, most of the health benefits end up being captured by the better off.

6.7 RAP Reforms: a review of their impact

In this section we review the evidence on the impact of the six RAP reforms components introduced in Chapter 4, with a particular focus on equity. In doing this we follow the flow of money within the system, and we go 'downstream', starting from the resource allocation and the priority setting components, and then turning to the analysis of the purchasing and payment system components, and finally the supply and demand side incentive mechanisms to improve coverage of essential health services (see Figure 4.1).

---

[160] For example, it may be impossible to enforce the provision of a basic package of services exclusively for the

6.7.1 Reforms of the Resource Allocation Criteria

According to Rice and Smith (2000), the main motivation driving those changes has been the achievement of tighter cost-control, but the new RA criteria have in fact also introduced a greater transparency and equity in the distribution of public funds. The capitation formula has been in several cases complemented to take into account the relative health needs of different segments of the population. Some services deemed essential were also often 'top-sliced' (assigned to national vertical programs before allocation of the rest to lower levels), and not included in the RA formula[161]. In some countries the RA formula has been corrected to take into account both individual (such as age and gender), as well as social and geographic factors positively related to need for health care services. For instance, the RA formula implemented in the '90s in Stockholm, Sweden, to distribute funds across different health authorities in different neighbourhoods, included age and gender of the resident population, housing tenure, marital and employment status, as well as previous utilization of health services in the computation of the capitation formula.

The experience from industrialized countries, such as the UK or Sweden, suggests that:

a. Population and need based funding creates a great opportunity to equilibrate distribution of resources across regions and districts.

b. Devising a gradual, smooth transition process to the new regime, so that the 'losers' in the redistribution process have the opportunity to adjust and the 'winners' to develop new planning capacity is a key but problematic issue. In this respect, in the UK, the Resource Allocation Working Party (RAWP) approach is widely perceived as a success. In the '70s a new capitation-based funding was adopted, which lead to a progressive rebalancing of the distribution of resources across the country, by assigning any real growth in the budget to under-resourced areas. Previously, the Southeast region and London had received a disproportionate share of the total. The transition period was set up so that if a region was over-resourced its budget could be cut by 2.5% maximum per year, whilst if a region was below target, it could receive an additional maximum funding equal to 5% per year.

---

poor, because doctors and other government health workers would not deliver it.

[161] For instance the share going to services belonging to the Basic Benefit Package (see following section), or that going to tertiary care and teaching hospitals.

Relatively few of the developing countries implemented reforms of their resource allocation criteria comparable to Sweden or the UK, while the majority of them continued to allocate public resources across regions, districts and health plans on a historical-incremental basis, with some adjustments through political negotiation. However, RA reforms were introduced in a number of countries in Latin America, Asia and even Africa.

In Latin America, countries that introduced significant changes in the resources allocation criteria, by introducing capitation-based financing, include Brazil, Bolivia, Chile, Colombia, and Mexico. [162] In all these countries, the variables that could be considered to adjust the capitation formula according to need were few, because of the lack of population-based data concerning individual and social circumstances as well as utilization of services. However, overall the impact of the new RA allocation mechanisms was positive and significant.

In Colombia, the government introduced a new RA mechanism based on a capitation formula for districts and municipalities, and adjusted it according to the share of the poor in each municipality and the fiscal effort at the local level. Brazil adopted a population and need-based formula to determine transfers to the states earmarked to finance primary care infrastructure, equipment and other capital expenditure (see following case study).

Case Study 6.3: Resource allocation criteria for capital expenditure in Brazil[163] (Musgrove, 1996)

In Brazil, the formula used to decide the extent of capital investment resources given to each state was specified as a weighted average of the state's share of the total population, and an indicator of the relative health need of the state vis-à-vis the country as a whole. The formula was specified as follows (Musgrove, 1996, p. 3):

$F_e/F = a(P_e/P) + (1-a)(X_e/X)$, where e refers to a particular government level, which, in the Brazilian example, was the state level (Estado); F is the investment, or other funds to be pre-allocated; P is population; X is a synthesis indicator of need for additional investment, beyond that associated with population alone ($\sum X_e = X$; $X_e \geq 0$;); a, where $0 \leq a \leq 1$, is the weight given to the population, as opposed to that assigned to the need indicator X_e.

---

[162] Baeza and Cabezas (1999) provide a throughout analysis of the experience with prospective funding and risk-adjusters in Latin America.

[163] Source: Musgrove 1996 HRO Series, World Bank, Washington, DC.

Initially X was computed simply as the inverse of total recurrent public expenditure in each state (labelled G_e for *'gasto'*), based on the assumption that where little was spent there was a greater need of further investment, and on the lack of data that could be used as alternative measures of need. However, using total recurrent public expenditure instead of per-capita expenditure, led to grossly overestimate 'empty' states' need. For example Amapa, which accounts for 0.21 percent of the total Brazilian population, received about eight times more than it would have received based on population alone. Later on, relative recurrent expenditure per capita replaced total expenditure in the computation of X_e.

The primary aim of the above prospective formula has been to achieve greater cost-control over local spending, to bring greater transparency, and to reduce inequities in capital investment across sub-national units. States such as Parana, which used to receive a disproportionate share of total capital expenditure had their share reduced to the benefit of the poorer states located in the North and Northeast.

In Chile, the government created a Municipal Common Fund (MCF), which is a horizontal equalization fund that receives up to 60 % of the wealthier municipalities' own-source income and redistributes it to the poorer municipalities on the basis of a per capita formula. The formula took into consideration the percentage of people living in rural areas, as well as the municipalities' capacity to generate their own revenue. As table 6.6 shows, the fund has significantly contributed to decreasing variability in per capita funding across different municipalities (the ratio of total public health expenditure per capita in the wealthiest to the poorest municipalities was reduced from 2.2 in 1991 to 1.6 in 1996 [Bossert 2000])

Table 6.6: The Impact of the Municipal Common Fund in Chile

	Income before MCF	Income after MCF
Average	24,646	40,823
Variation coefficient	17,437	30,984
Gini coefficient	0.45	0.30

Source: *Bossert, 2000b*

Chile also provides an interesting example of how the capitation formula can be complemented by other funding mechanisms to support specific initiatives at the local level and to achieve specific pro-poor impact. The government utilized Matching Grants, piloted in the District of Nuble (Bossert, 2000b), to channel local funds towards government priorities[164]. In Nuble the Matching

---

[164] With a matching grant, the agreement between the central and the local government is the following: if the

Grant instrument was used to finance a program intended to fight poverty in the area by providing new services of public utility, education, as well as health services. The central government covered about 60 percent of the total costs of the project, and the local government the rest 40 percent. Bossert (2000b) reports that the impact on the poor was positive.

Resource allocation reforms have also been undertaken in South Africa, Uganda, and Zambia. In these countries, the implementation of the new resource allocation formula was not without difficulties, and progress has been uneven. In Zambia a new population-based formula was implemented in 1994, with some crude adjustments to reflect cost and need indicators. The absence of a fuel station or a bank was used as a proxy for underdevelopment, and districts received an extra weight where these facilities were missing. Subsequent research (Lake, Mtonga, and Nakamba 2002) showed that the correlation between these first crude need indicators and more sophisticated measures of relative deprivation (derived through principal components analysis) was significant, and that other simple indicators such as population density and remoteness could explain most of the provinces' relative deprivation. One key issue that hindered the impact of the new population and need-based formula in Zambia was that a large share of the total public health expenditure was "top-sliced" to fund tertiary care facilities, concentrated mainly in Lusaka.

In South Africa, which before 1994 had one of the more inequitable health systems in terms of resource allocation (McIntyre and others 2000), in the immediate post-Apartheid period there was a strong political commitment in favour of a radical shift of resources away from the relatively prosperous areas and toward the poorer provinces, mainly in the Northern and Eastern parts of the country. The allocation to the poorest Northern province was expected to increase from between 6 and 7 percent of the total share to 15 percent over a five-year period, while funds for Gauteng province were to decline from 25 percent to between 17 and 18 percent, a significant reduction in real terms. Moreover, 30 percent of this shift was expected to occur in the first year. Not surprisingly the above targets proved impossible to achieve: Gauteng received extra support in the short term and the Northern Province was unable to absorb the additional funds effectively. Also, the devolution of powers to the provincial governments (which started to receive a block grant without specific allocation to health, and had different spending priorities) and the new macroeconomic policies (which were characterised by restrictive fiscal policies in the 1990s) hampered the plan to increase

---

latter endorses a specific program proposed by the former, and it is willing to contribute a portion of the total cost of the new program, the central government assumes the responsibility to cover the rest.

expenditure in health care and other social services in the poorer provinces (Gilson and others 1999, McIntyre and Gilson, 2002).

This brief presentation of the RA reforms indicates that the new population and need-based RA funding formula are capable of achieving a strong, positive impact on geographical equity (Bossert, 2000 and Pearson 2002).It also suggests that there is a great potential for further expanding population and need-based RA formula, and for using more articulated risk-adjusters to pursue equity priorities. As a first step to develop RA formula, up-to-date demographic census data would suffice. In fact, the experience in Zambia (Lake et al., 2002) shows that significant progress can be made simply with reasonable census data on population and age/sex breakdowns, plus some proxy indicators of deprivation. Then, a need-based RA formula would require information on the following variables[165]:

1. Standardized Mortality Rates;
2. Morbidity measures;
3. Share of population in poverty;
4. The amount of other funding sources, for instance of donors' expenditure in each sub national unit.

Data on the above variables are at least partially available in middle-income countries, but are rarely, if ever, present in low income countries. Thus, more sophisticated formulae can be more sensitive to health needs, but they are too demanding for low income countries[166].

The existing experiences suggest that a number of preconditions must be met in order to support the implementation process of the new population and need based RA criteria:

• Government's funding must be a non-negligible part of total health funding, and the government can decide how to allocate geographically.

• Supporting changes and improvements in overall planning procedures need to take place, particularly those related to personnel and capital planning and management[167]. Adequate information systems must be developed to guarantee access to timely and reliable data.

• There must be basic capacity to regulate financial flows - i.e. to ensure the resources get to the intended destination, and reach the intended beneficiaries.

---

[165] It is important that all the variables used in the RA formula cannot be influenced by the level of government receiving the funding.
[166] RA formulae can and should then be updated as necessary and when new data become available.
[167] The RA formulas should be reflected in a 'medium term expenditure framework' and in national infrastructure development strategies, where they exist, and in long term planning documents.

Many of the above conditions may not be feasible in the poorest, capacity constrained countries, characterized by $3-4 per capita public expenditure for health, and 80-90 percent of total resources for health care spent on an out-of-pocket basis. For example, consider the situation in several of the countries of Sub-Saharan Africa, where actual releases bear little relation to budgets, flow of funds is erratic, doctors, nurses and other key 'productive factors' are in fact engaged in different activities and away from where the financial flows records suggest, and most of the resources going to the health sector are from patients OOP directs payments. In such a context new RA mechanisms would not be able to steer incentives towards more effective and equitable objectives.

6.7.2 The Priority Setting Component of RAP Reforms

Several among low and medium income countries during the decade of the 1990s started to define an essential health package (BBP, Basic Benefit Package, or ESP, Essential Services' Package), heavily influenced by the methodologies set out in the 1993 World Development Report (WDR). However, in the majority of cases the implementation of the BBP has never passed the preparatory phase, or it resulted in some tokenistic change under international donors' pressure.[168]

The few existing experiences where BBP-based policies have been really put in place indicate mixed results. In several cases essential packages have proven to be vaguely defined, setting out a range of services but giving little detail on who is to provide them, how they are to be delivered, how many services are to be delivered and to whom. For example, in 1995 the Government of Georgia introduced a new rationing policy inspired by the BBP (see Chapter 8). However, the range of services included in the BBP (originally only 6 programs) was quickly expanded under political pressure. The criteria for inclusion in the list of beneficiaries in the program for vulnerable groups (on the 6 programs and the only one where rationing was supposed to be individual and not service-based) also became extremely arbitrary. The Government of Georgia plans were never enforced, and because of the fiscal collapse of the country informal payments became the prevalent source of funding for health services, both for BBP and non-BBP services (see Chapter 8).

---

[168] Bobadilla (1996) presents a review of several National preparatory studies for the definition of an essential package of services, based on the analysis of the burden of disease and on the availability of cost-effective interventions. He reports that very few of those studies became translated into concrete proposals for

In theory, the reprioritization of government subsidies to health care should produce a positive impact from an equity perspective, increasing the share of the benefits that went to the poor (see section 6.3). However, in practice, in designing prioritisation policies countries were inspired largely by the cost-effectiveness principle, under the influence of the WDR 1993, which does not at all guarantee that health services which address the health needs of the poor with priority, as we indicated previously (see Annex 4.1, Chapter 4).[169]

In West European countries and the USA, in spite of a wide debate on the necessity of rationing or reprioritizing public health services because of limited budgets and raising expectations, so far only relatively marginal services have been excluded from government financing, such as some mental health and rehabilitation services. More significant results and savings have been achieved only in containing pharmaceutical expenditure, by defining different categories of drugs subject to different reimbursement policies. The experience accumulated so far in several countries and regions (among which the more well known are those in the Medicaid Program in Oregon, in Oregon, USA, in Sweden, and more recently, the UK) also indicates that exclusively expert-driven, radical reprioritization processes are unlikely to be accepted and to succeed, and that articulating new rationing policies is a complex political process that is likely to succeed only if it done gradually, involving all major stakeholders from the outset. Prioritization policies have been more successful when they were based on existing patterns of expenditure, and proposed realistic changes at the margin.

6.7.3 Purchasing/contracting reforms

In discussing the impact of purchasing reforms, it is useful to distinguish the cases where the purchaser is: a) a public agency (a social insurance agency or a sickness fund); b) a group of general practitioners; c) a private insurer.

Public Agency (Social Insurance agency or Sickness Fund)

Several countries characterised by universal health insurance, or striving to achieve it, in implementing the purchaser-provider split, entrusted new social insurance or sickness funds with the

reprioritization of government services, and even fewer of these proposals were actually implemented.
[169] When the focus is shifted from services to people, cost-effectiveness and equity motivated reprioritization policies do not lead to equal results, because those that are (marginally) more able to benefit (and should then come first according to cost-effectiveness) are generally the wealthier and/or those already in better health.

purchaser role (see Chapter 4). These countries included the UK (GP Fund holding scheme), the Czech Republic, Poland, Georgia, Chile and Colombia in Latin America. The "public" purchasers were partially separated from government, and in some case were allowed to directly compete for patients (for example in the Czech Republic).

The evidence available indicates that reforms achieved mixed results. On the one hand, they contributed to build new managerial capacity, produced some competition, and possibly greater quality and more transparency (see below). On the other hand, the need to allocate money to sustain the newly created funds diverted scarce resources from other more urgent and productive utilization. In addition, where competition was introduced, there is indication of cream-skimming of patients. To limit these effects, reforms plans to introduce competition among several health funds and insurance companies in countries traditionally characterised by universal social insurance such as Germany and the Netherlands, foresaw the creation of a compensatory fund to discourage risk-selection. Unfortunately, only demographic criteria such as age, gender and residence, which are poor predictors of future health services' utilization, can be used as risk-adjusters to cross-subsidize premiums (see Chapter 2).

In addition, lower income countries still exhibits legal "limbo" in the regulation of health insurance (for instance, the absence of a clear bankruptcy law). These limbos, coupled with weaker governance and regulatory capacity, may create a situation where financially irresponsible conducts and bad management could thrive, as the experience of the Czech Republic demonstrates (se case study below).

Case Study 6.4: Competition on the purchaser side? The case of the Czech Republic, 1991-2000

In 1991 in Czechoslovakia a new Law, the General Health Insurance Act (ACT 550), shifted health financing from the general budget to a payroll tax set equal to 13.5 percent of gross salaries, separated purchasers and providers, and create competition on the purchaser side. By 1993, 27 new private insurance funds were active in the health insurance market, alongside the National Health Insurance (VZP) Company. When the reform process was initiated, no explicit discipline for bankruptcy existed and providers were reimbursed according to a fee-for-service payment system (a cap had originally been imposed on total expenditure, but it was rapidly abandoned under providers' pressure). As a result of the reforms, between 1992 and 1998 per capita health expenditure grew in real terms from 5.3 to 7.2 percent of GDP (to 9.2 percent, by also taking into account sickness allowances and community level public health programs). In 1995 the first of the newly created insurance funds went bankrupt, followed by others. Moreover, as a result of new government regulation

(setting a minimal number of enrolees, a maximum level of administrative costs etc.) other funds decided to merge. By 1998 only 9 insurance funds remained out of the initial 27. In 1995 the National Health Insurance Company (VZP) enrolled only 64 percent of the total population. By 1998, the proportion had risen to 80 per cent. In the last few years, the extent of public regulation of the insurance market has been substantially expanded. It now rigidly disciplines premiums and benefits, and several other aspects. Under such a tight regulation, "Insurance funds have the incentive to be more "passive payers" rather than active buyers (Maly, 1999, p. 4)".

By contrast, where competition among purchasers was not allowed (and only one regional fund was created covering the entire population in each district or region), the purchaser reforms did not lead to any significant change, or to a rationalization of the delivery system. Many factors contributed to this outcome.

➤ In some countries, for example in Poland or Georgia, the overlapping of responsibilities among the insurance funds, the Ministry of Health, the other Ministries (Finance, Education) and the local governments created an uncertain and convoluted governance structure. In some cases (ex, in Poland) the government assigned the sickness funds the role of financing providers, but maintained or devolved to local governments the ownership of hospitals and of polyclinics. The prerogatives of the "funder"(the insurance fund) as opposed to those of the "owner" (the central or local government) were never precisely defined. As a result of the convoluted governing structure created by the reforms, providers and particularly hospitals in practice became non-accountable to anyone (Jakab and Preker, 1999). In the end, in some case the sickness funds were "reabsorbed" by the central Ministry of Health, or acted as simple "cashers" of the system rather than real purchasers.

➤ Moreover, the health financing structure remained very fragmented, both in terms of sources of revenue and in terms of power to allocate funds. Sickness funds, financed from payroll taxes and social protection contributions, never controlled the majority of sources of revenue for health facilities, which remained OOP payments by patients and direct government financing. Health care providers that did not receive enough financing -through contracts and service agreements- from the sickness funds found a way to be bailed out by the central or local governments Allocation of capital expenditure also remained under central governments' control.

➤ Finally, absent competitive pressures and accountability mechanisms, in several countries the new insurance funds could become as monolithic and consumer-unresponsive as the bureaucratic apparatus that they were meant to replace.

Primary Care doctors as purchasers

In several countries during the decade of the 1990s a new specialty in family medicine was created or strengthened, in an attempt to provide more doctors to rural and underserved areas and or to reduce costs by restricting access to specialist and higher care. In the latter case, for example in the UK or Poland, General Practitioners (GPs) were assigned a gate-keeping role for access to subsequent levels of care. Some countries went further and assigned GPs (capitation-based) budgets, which they could use them to refer patients to secondary and tertiary care facilities. GPs started to act as purchasers of services, and set contracts with service providers. According to the evidence available, they were much more innovative purchasers than sickness funds, and lead to improved quality of care for their patients (see case study 4.2 in Chapter 4). However, the "purchasing" role of GPs has also received severe criticism. The claim is that GP budget-holders under-treated patients (see case study below on Poland), and that phenomena of cream skimming and risk-selection increased (see first trade-off discussed in Chapter 4). In the 1990s in the USA similar criticism was common against Managed Care Organization, which made use of GPs as gate-keepers for specialist care.

Case study 6.5: The Polish experience with primary care reform

From January 1999, in Poland each individual primary-care doctor receives an adjusted capitation funding and she/he is financially responsible for referrals of patients to most outpatient specialists. At the same time, it is compulsory for patients to obtain a referral from their primary care doctor to access specialist care. A similar organizational model has been implemented in 1991 in the UK (see Chapter 4). However, while in the UK only groups of doctors having at least a minimum number of patients (5,000) were assigned a budget and financial responsibility for referrals, in Poland the scheme was introduced at the individual level. In the absence of risk-pooling mechanisms, the new referral system creates a strong incentive for primary care doctors to under-treat and dump potentially costly patients. Moreover, the only way patients can bypass the cumbersome referral system is through emergency admissions. Indeed, in 2000 there was a rapid increase in hospital admissions (30-35% increase in one year), immediately after the implementation of the reforms. Tighter regulation and limitations to the referral system were subsequently imposed.

Private Health Insurers as Purchasers

In some countries where health financing and provision have traditionally been private, for profit health insurers have progressively taken on the role of purchasers, mainly in order to achieve a better control over providers and reduce costs. The "purchaser" role emerged either out of a spontaneous

evolution in the market, or under the pressure of regulatory reforms (see Chapter 4). These developments only marginally affected the potential negative effects of competition in the health insurance market, which we have extensively analysed in Chapter 2. As we have already seen in discussing reforms in Western Europe and Chile (see Chapter 2), the conclusions of the theoretical literature seems to be of practical relevance; the evidence shows that competition on the purchaser side induces adverse selection and, in the presence of a public insurer of last resort, cream-skimming (see discussion of first trade-off in Chapter 4).

Contracting

With the purchaser-provider split, explicit contractual agreements with private or "corporatised" - not-for-profit or for profit- providers have become common practice (see Chapters 4 and 5, for an explanation of the term "contract"). The three forms of contracts more widely utilized have been the following (Loevinsohn, 2001):

- Contracting Out (CO): in which the contractors assumes complete responsibility for service delivery, including hiring, firing and setting wages, procuring and distributing essential drugs and supplies.

- Contracting In (CI): where the external contractors are brought in to provide certain specific productions inputs, but always within the umbrella of the public health system. NGOs may for example be asked to manage the health centres in a particular district, using the logistics, drugs and supplies provided by the government. The contractors cannot hire or fire health workers, and the use of any surplus is strictly regulated (usually it must be used to improve quality of services).

- "Leasing" where an organization offers to manage and finance existing health facilities (e.g. an NGO offers to take over and manage poorly performing health centres).

Still very few rigorous studies exist to assess the overall performance of the new purchasing and/or contracting arrangements, particularly their impact on equity, and most of the evidence available is from non-experimental studies (with no control group) (for Latin America, see the review of the evidence by Slack and Savedorf, 2001; for Cambodia, see Mills, 2000). A review by Mills and Bloomberg (1998) pointed out that the literature has concentrated more on the promotion of

contracting policies than on evaluating their results first. Furthermore, performance evaluations are often elaborated by the same contractors who are involved in the contract design or implementation.

In synthesis, the available evidence reveals both promising, as well as problematic aspects of the new contracting experiences. A study from Cambodia (Bhushan and others, 2002) suggests that contracting in and out services to NGOs produced an extremely positive impact on the poor, improving coverage and quality of services for lower-socio-economic groups, as the following Table 6.7 indicates. The benefits were greater for contracted out districts, where government service delivery was more expensive ($4.50 per capita, compared with $ 2.82 and $1.5 spent per capita in contracted in and control districts), but where private out-of pocket payments were reduced by 27 percent overall and over 70 percent for the lower socio-economic groups, indicating good targeting of beneficiaries.

Table 6.7. Average Change in Service Coverage Indicators in Cambodia (first 2 and ½ years of reforms)

Indicator	Control District	Contracted-in	Contracted out
Immunization rates-all[170]	56	82	158
Vitamin A capsule Receipt –all	-25.1	18.1	20.9
Vitamin A capsule Receipt –lower 50% socio-economic segment of population	-24.1	29.9	23.9
Percent of illnesses Treated in Public Health Facility—lower 50% socio-economic segment of population	81.7	490.5	1096.0

Source: Bhushan and others (2002)

Utilization of curative services in district hospitals by the bottom 50 percent socio-economic group increased about twelve fold in contracted out districts, six fold in contracted-in districts and only less than double.

The same positive results are reported in local studies which analyzed contracting with not-for-profit providers (NFPP) in Bolivia, Bangladesh (nutrition services), and India (TB-DOTs treatment). The flexibility in the management of inputs allowed NGOs to improve service coverage, and achieve

---

[170] Baseline immunization coverage was 25.5 percent in contracted out districts (COD), 29.9 in contracted in districts (CID), and 34 percent in the control districts (CD). The follow-up survey showed coverage of,

significant cost savings compared to government standards without compromising quality of services. A World Bank analysis by Loevinsohn and Harding (2004) reveals that in India and Pakistan, non-governmental entities achieved better coverage of essential services than government facilities with similar financial resources. A study of the impact of contracts in South Africa (Palmer 2001) indicated the importance of maintaining competition among providers, and of defining ex-ante the population to be covered by the contract.

By contrast, several other studies have pointed at the significant human capital and information systems requirements to specify and to manage contracts effectively (high transaction costs), to be contrasted with the limitations characterizing several developing countries (McPake & Banda, 1994, Mills et al., 1997, Palmer, 2001, Taylor, 2000, Van Holten, 1998, Vining and Globerman, 1999, Waters et al. 2002). From the evidence available, one can say that effectively designing and monitoring contracts has proved as demanding as directly managing health services, and that where the public sector was not able to effectively 'row', it has generally also proved unable to effectively 'steer' independent or semi-independent providers through contracts. Because of lack of funds and managerial capacity, absence of competition, and in some cases widespread corruption, contracted out services can perform more poorly than government services (Liu et al, 2004). For example, the US Medicaid system suffered from different kinds of internal fraud and abuse due to over-billing, treatment splitting and other illicit actions by the private practitioners. In Colombia, the contracting-out of the administration of the public subsidy for a Basic Benefits Package resulted in the registration of many non-existent persons or double-registration thus skimming off public health funds for private gains. Countries in which corruption is pervasive will be facing particular challenges in this aspect[171].

The following case study from Bangladesh illustrates some of the problems encountered in contract management.

Case study 6.6: Bangladesh experience with engaging the not-for profit sector in delivering nutritional services [172]

In Bangladesh, there is an ongoing experience with contracting with NGOs for nutritional and primary care services in two large projects, the Bangladesh Integrated Nutrition Project (BINP) and the Urban Primary

---

respectively, 65.8 (COD), 54.4 (CID), and 53 (CD).

[171] For example, in many poor countries, public doctors also work in private practices– often during their official working hours, and therefore contracting-out of health services to the private sector could lead them to work even less in the public sector during their official hours.

[172]Source: Loevinsohn (2001).

Health Care Project (UPHCP). BINP, for example, has contracted with 7 NGOs to provide nutrition related services to a population of about 10 million and has been in operation since 1997. Loevinsohn (2001, p.1 and subsequent pages) reports: "Contract management has been a significant issue under BINP with one NGO owed more than $3 million and the other contractors also owed substantial amounts. There have also been significant problems with the Government's procurement and distribution of supplies and equipment, particularly in comparison with those materials procured by the NGOs, which have been readily available. Field supervision and monitoring by BINP has been problematic. The mid-term review of BINP examined the change in nutritional status in 6 project *thanas* and 2 control *thanas* as judged by a third party. It appears that the decline in malnutrition in project *thanas* was, at best, only modestly greater than in non-project *thanas*".

In conclusion, despite the existence of successful experiences in contracting, particularly in terms of improvements in access to health care (see below), valid criticisms persist on the possible impacts of such policies, for example in terms of health system equity. Hence, different authors refer to the need to generate institutional capacity for negotiating, implementing and monitoring the contracts (Mills and Bloomberg, 1998). The experience from industrialized Nations seem to suggest that over time contractual relations evolve into long-term relations based on trust; this in turns leads to a decline of competitive tenders, and the decreasing importance of prices after the first round of contracting. In other words, initial contracting is competitive, but renegotiation is not. Information from the USA and the United Kingdom points in the same direction.

6.7.4 Payment systems' reform

This has been one of the reform components that received most attention, and it was assigned a high priority in a number of high and middle income countries over the last twenty years. In the first part of the '90s, several countries (including Italy, Hungary, Czech Republic, Croatia, Brazil, Colombia, Thailand and several others) moved from input-based towards case-based PS for hospitals. Reformers claimed that the new cost-per-case or fixed price payment system would lead to greater efficiency by subjecting providers to a hard budget constraint, and reducing funding for scarcely utilized facilities, and would give the incentives for greater patient-responsiveness by implementing the principle that "money follows the patient".

The evidence available on PS reforms is incomplete and not always rigorous, and it is affected by the same limitations discussed for the purchasing and contracting component of the RAP reforms. In addition, there is almost no information available on the impact of the new PS on the distribution of service utilization by different socio-economic groups (impact on equity). Nonetheless, the evidence

we have on PS reforms suggests that there is no such thing as a unique "ideal" PS, and that "institutions matter". Thus, the empirical evidence on PS points at the existence of trade-offs, confirming the theoretical results analysed in Chapter 5. Moreover, an adequate institutional capacity and a well-designed and functioning legal and regulatory framework are crucial in determining the performance of any PS, and in limiting its potential adverse effects.

Let us briefly summarize here some of our previous theoretical results that are more relevant for the analysis of payment systems' reforms, focusing on the incentives associated with cost-per-case or fixed-price payment systems. If we denote by Π the hospital surplus, by x the number of cases treated, by p the unit or average payment that the hospital receives, and by c hospital costs, $\Pi(x) = p(x)x - c(x)$, a cost-per-case reimbursement is equivalent to fixing ex-ante the unit price or average revenue p. Under this PS, the hospital can increase its surplus either by reducing costs, by treating more patients (until price is equal to marginal costs), or by treating more profitable cases, and minimizing resources spent on each case treated. Thus, this payment system can either result in greater productive efficiency (as in the yardstick competition model, see Chapter 5) and more patient admissions, or lower quality and in the extreme case, dumping of patients. In addition, providers subject to cost-per-case PS, could try to misreport the existing case mix to increase the payment received per unit of service (see model in Chapter 5).

In Former Socialist Countries, where hospital managers had extremely weak control over unit costs (mainly because of National laws rigidly disciplining deployment of labour and capital), the new fixed-price payments systems generated some efficiency gains such as a reduction in average length of stay, but in general they did not contain hospital costs, and lead to significant increases in hospital admissions (see following Case Studies).

Case Study 6.7: Reform of hospital payment system in Hungary

In 1993 Hungary introduced a new payment system for hospitals, which at the time was the most sophisticated one in the region, based on the number of discharges according to the DRG classification (Homogen Betegseg Csoportok in Hungarian). The new payment system was continuously refined for almost a decade. Despite this extensive effort, the new payment system did not generate the results that were expected. There was not any significant change in the trend of hospital admissions, which rose from 21.8 to 24.2 per 1000 between 1990 and 1996 (see Figure 6.); average length of stay decreased from 9.9 to 8 days (a positive result, but in line with previous trends), and the share of hospital expenditure over the total remained roughly constant. The national case-mix index increased from 0.97 to 1.10 (Dorotinsky, 1998; Orosz and Hollo, 1999). Local administrators

increased revenues under the new payment system by increasing hospitalisation and by misclassifying patients (DRG creep). Finally, between 1991 and 1997, employment in the health sector decreased by only 2 percent.

Figure 6.7: Hungary: Impact of Payment Reform on Hospital Sector
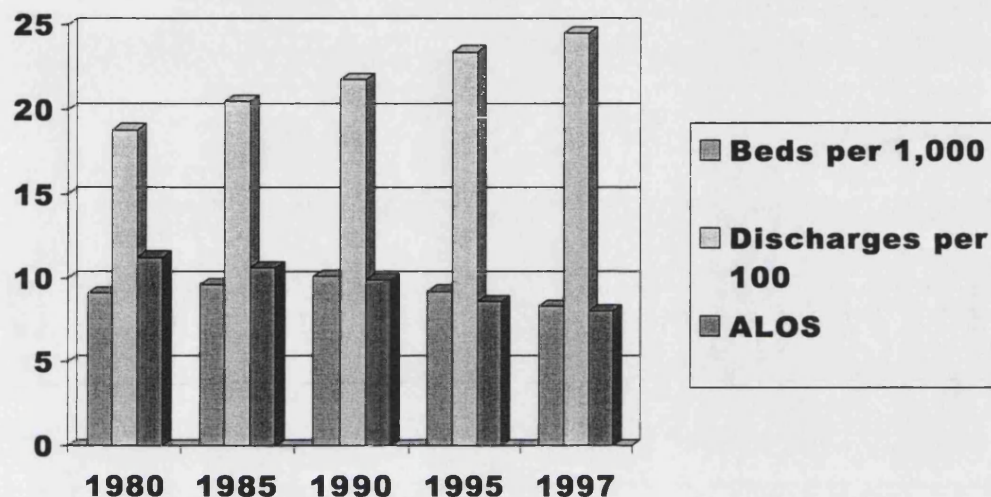
Source: Langenbrunner & Wiley, 1999

6.7.5 Supply side Incentives in Pro-Poor Health Programmes

As explained in Chapter 4 (sections 4.4.5 and 4.4.6), in some countries RAP reforms introduced explicit financial and non financial incentives for providers, patients or both to increase utilization by the poor or of specific essential health services primarily consumed by the poor. In Case study 6.7, we review the experience on the use of enabler and incentive schemes to improve coverage and outcome of TB treatment, based on Direct Observed Treatment (Weil et al., 2001).

Case study 6.8: The Use of Provider Incentives to Improve TB Program Performance[173]

Tuberculosis (TB) remains a major killer of adults worldwide, only surpassed by HIV/AIDS, with two million persons dying due to TB last year and the toll worsening rapidly in some regions, especially Sub-Saharan

---

[173] The case study is an extract from work by A. Beith, R. Eichler, J. Sanderson, D. Weil.

Africa with the intimate linking of the two epidemics. The poor are hardest hit by TB, given the nature of transmission of infection and breakdown to disease, but all income strata are at risk. A proven cost-effective strategy to detect and cure TB that is capable of reducing mortality and morbidity even in very poor settings is known as Directly Observed Treatment Strategy (DOTS). Among the top challenges faced by DOTS programs throughout the world is ensuring that patients are able to initiate treatment quickly, that they stay motivated to complete the full course of therapy, and that providers assist patients in this process, as well as provide safe and appropriate care.

Recognizing the need to improve DOTS-based TB control programs, several projects have introduced schemes that provide incentives to enable and motivate providers and/or patients to pursue DOTS objectives. Some incentives are not linked explicitly to performance, but others relate directly to measures of improved performance. Examples of the former include providing improved training opportunities for TB workers or per diems for service supervision; examples of the latter include providing financial bonuses to providers for each infectious patient that completes treatment, or food packages to patients that present regularly to take their medicine. Some of the most successful DOTS programs worldwide, such as in China, Peru and Cambodia, use performance-linked enablers and incentives. However, the evidence collected so far on these programs is largely incomplete. A preliminary survey of the literature on enabler and incentive schemes linked to TB control projects indicated that few schemes involved rigorous design, monitoring or evaluation elements, or had clear financing and management plans. No evaluation studies focused on incentives in contractual relationships between governments and private, for profit or non-profit, providers, and no studies attempted to measure the cost-effectiveness of different schemes (plus, among the existing 23 publications only 2 were from developing countries). Design, implementation, and impact of incentives schemes for TB treatment seemed to vary significantly[174] , but that overall they did influence behaviour and lead to improvements in program performance. The survey of the existing experiences also showed that there was considerably more concern about formal provider schemes than patient-focused schemes. This apparently related to the strong perceived potential unintended negative effects on performance in other areas or for future motivation, if resources for incentives are no longer available.

Some of the more significant conclusions reached so far are summarized in the following:
   i.   Problems in motivating providers and enabling patients are important and relevant to effective TB control throughout the world.
   ii. These problems may especially inhibit access and utilization by poor patients or participation by providers working under difficult conditions.

---

[174] Some schemes are clearly defined and tied to measurable actions or results while others are less regularized. Some schemes depend on sporadic funding while others are fully incorporated into their program's annual budget. Some are offered by the public sector and others through NGO or private providers.

iii. Incentives exist for all stakeholders in TB control – whether they are explicit or implicit. Some of these incentives may inhibit good performance and are not aligned with the objectives of TB control.

iv. Even with DOTS implementation, some functional constraints inhibit patients and providers from participating fully and effectively. Formal incentives that encourage actions to be in line with DOTS norms may improve performance.

v. Enablers and incentives may be appropriate and help improve coverage and cure rates in DOTS programs, which have already achieved cure rates of 70% or more. They may not be right to consider if fundamental problems of standards, inputs, training and supervision have not yet been at least partially alleviated.

vi. Appropriate enablers and incentives may vary considerably depending on underlying epidemiological, economic and social conditions, health system structures and capacity. There is likely a continuum of approaches that may be effective from specific local measures to standardized national approaches.

There are potentially serious unintended negative effects of enablers and incentives that must be avoided through proper design and monitoring.

In some countries Governments targeted the subsidies to improve service coverage for the poor to not-for-profit providers (NFPP). For example, in Uganda NFP hospitals run by religious organizations have traditionally been close to poor and marginalized groups service beneficiaries, but during the 1980s and 1990s they had to significantly raise user-fees, due to financial difficulties (raising costs of provision and reduced donations). In 2000 the Government of Uganda decided to offer subsidies in exchange for agreements on service provision and price controls. These subsidies did allow NFP hospitals to reduce user-fees, and created a positive impact on utilization (see Giusti, 2000).

In summary, the evidence of the impact of the supply-side incentives is still largely inconclusive, but from the evidence that we have it seems that they are increasing utilization of services by the poor.

6.7.6 Demand side Incentives in Pro-Poor Health Programmes

In other countries subsidies were directly given to the poor and marginalized groups. For example, Gortler et al. 1999 and Armstrong et al., 2002 investigated the impact of vouchers for family planning and basic care given to sex workers in Nicaragua, and for mosquito nest in Tanzania. Both programs were not without difficulties, but they overall lead to significant increases in utilization and improved health outcomes. Case study 6.8 reviews the experience in Nicaragua.

Case study 6.9: Vouchers for reproductive health services in Nicaragua (Source: Gorter et al. 1999)

Since 1995 a voucher programme has been operating in Managua, Nicaragua, with the aim of increasing the uptake of reproductive health services by female sex workers. Every 3-5 months approximately 1200 vouchers (corresponding to the estimated number of sex workers operating at any given time in the city) are distributed by fieldworkers and NGOs at prostitution sites. The vouchers entitle the sex workers to free services at one of 8-10 private, NGO and public clinics, which are contracted to the voucher agency by competitive tender. Approved providers must follow a set treatment protocol, and receive training. Contracts are reviewed after each round of voucher distribution, and renewed subject to an assessment of quality of care. The clinics return the vouchers to the voucher agency, which reimburses the provider an agreed fee per voucher. The sex workers were involved in the design of the programme.

In the evaluation study, in each round of the survey 10% of recipients are interviewed about their experience. Initially, sex workers reported that the gatekeepers to care (nurses and receptionists) lacked sensitivity: training. Sensitization of this group helped to improve their attitudes towards these clients.

Technical quality of care was also low, and training and treatment protocols were introduced.

While the prevalence of STD is only slightly lower than at the beginning of the project (possibly due to a high turnover of female sex workers), incidence of STD among women who have used vouchers more than once dropped by 65% in the first three years of the programme. Following a recommendation by the sex workers, they now receive vouchers to give to their regular partners and/or clients as well. Sex workers appreciated the fact that they could choose which clinic to attend, and made their choice on the basis of distance and friendliness. The clinics reported that their main benefit was improvement in the technical quality of their services, and that the lessons learned were applied to all of their clients. They felt that their reputation was enhanced by being contracted by a prestigious public health agency (the Central American Health Institute).

6.8 Conclusion

In Chapter 4 we described the principal elements of RAP reforms, and we explained that these different elements were likely to produce a distinct impact on the performance of the health system in terms of efficiency, equity and quality of services. In this Chapter the focus has been on equity. We initially presented the evidence on current distribution of health and health care utilization across countries and within each country, which shows striking inequalities between high and low income Nations, and between rich and poor in low and middle income countries. Then, we reviewed the different equity principles that have been proposed in the literature. Regardless of the notion of

equity in health that we adopt, the distribution of health outcomes and health care benefits in the majority of poor and middle income countries can only be described as abysmally inequitable. By contrast, the equity assessment of ongoing RAP reforms is more complicated, particularly when these reforms have improved coverage of the poor, but have also increased disparities across different socio-economic groups (see case study 6.2; see also discussion of Figure 4.5 in Chapter 4).

In the second part of the Chapter, we have reviewed the empirical literature on the impact of the RAP reforms, with a specific focus on equity. We repeatedly underlined that the evidence on existing reforms is incomplete, and that ongoing and future reforms need to be subject to more rigorous monitoring and evaluation. It is unacceptable that radical health systems reforms -not only in SS Africa or South Asia, but also in Central Europe or West Europe- which potentially affected the lives of millions of people, have been implemented without any systematic monitoring and without rigorous collection of baseline and follow-up information, or a control group, which could allow an ex-post rigorous evaluation of their impact. The equity dimension of the reforms has been particularly neglected, with hardly any information of the changes in utilization by different socio-economic groups (a notably exception is Colombia; see Case Study 4.1).

On the basis of the evidence which is available, it is possible to say that the new resource allocation formula and the supply and demand subsidies for pro poor programs in general had a positive impact on equity, by making the distribution of resources across different parts of any country and among different socio-economic groups more transparent, more equal, and in the best cases more need-based, and by reorienting public expenditure on services that disproportionably benefit the poor. By contrast, the reprioritization reforms were largely irrelevant, and the purchasing and payment system components of RAP reforms had either a negligible or a negative impact on equity. Specifically, purchasing, contracting and payment system reforms meant to reduce costs, increase efficiency-or improve quality of services may in fact have lead to widen quality differentials across services, and increase providers' incentive to cream skim, risk select and under-treat poorer or more vulnerable patients. However, in the absence of any solid empirical evidence, the arguments in favour or against remain primarily theoretical.

Annex 6.1 According to which criteria should resources for health care be distributed among individuals?

What does it mean that a certain distribution of health, or health care services, or that a specific repartition of the burden for financing health services are equitable or inequitable? In this Annex we chose to reconsider the notion of equity starting from basic principles. We believe that attempting to present this discussion by broadening our perspective to other sectors and by going back to basic principles is useful, for three reasons: a) to provide the means to read with greater awareness the assumptions at the basis of the literature on equity in the health sector, equity and reforms, etc.. b) To propose in a unitary, synthetic, and hopefully original way several among the most significant contributions to the matter of equity and health care produced in the last two decades, which is a body of literature not many are familiar with. c) To frame the issue of equity in health care within the government versus markets debate. The main points of the discussion are presented here, whilst further details are presented in Annexes 6.2 and 6.3.

Since the literature that has analyzed this question specifically focusing on health and health care [175] draws upon an immense body of literature in economics[176], sociology, political science and philosophy[177], we selectively revisited this body of literature.

The key question we try to address is the following: "According to which criteria should resources for health care be distributed among individuals?" First, it is useful to clarify the distinction between equity and equality. As Le Grand (1991) correctly points out, the two concepts should be kept distinct. Equity, like justice, is a normative concept, whereas equality is an empirical, descriptive

---

[175] Among all the articles and books on the issue of health and equity, I believe two contributions stand out for clarity and thoughtfulness, and are thus recommended to those interested in further reading: Le Grand (1991) and Culyer and Wagstaff (1993).

[176] Le Grand (1991, Chapter 5) provides a clear and insightful review of the main criteria advocated in economics, equity as envy-freeness, as equality of income, of utility, or of marginal utility, and as absence of rank-reversal.

[177] In modern times what characterizes a just or equitable society and a just distribution of resources, assets, etc. has been a matter of contention since the early 17th century, when the Iusnaturalist philosophers put equity at the centre stage of the philosophical debate. They were so named because they believed in the existence of a "Ius Naturalis" a Natural Law, which preceded and was superior to any earthy power. According to such Natural Law men were created with some equal non-alienable rights. John Locke (1634-1703) was one of the first and more influential Iusnaturalist philosophers. All modern positions based on the notion of human rights find their root in the Ius Naturalis philosophy.

concept. In general, achieving greater equity does not necessarily imply moving towards more equality; the opposite may well be true[178].

Once the distinction between equity and inequality is clear, the answer to the above question stems directly from the analysis of how unequal the distribution of resources for health care is in our societies, compared to the "benchmark distribution" that we regard as just. A particular distribution is considered inequitable is there is discrepancy between the two and if its causes are amendable.

Hence, the first question to ask is the following: "What would be a just distribution for resources for health care? The answer to this question will depend on whether or not we believe that health is a fundamental human right (for a further discussion of this point, see Annex 6.2).

In general, the benchmark distribution for any given resource or asset is not perfectly egalitarian, and it can be different according to each of us. However, for everything that is considered as a human right, matters are different and more straightforward. By definition, individual human rights, unlike assets and resources, are inviolable, inalienable (one cannot buy or sell his/her right to vote, or his/her right to free speech) and must be equally enjoyed by all. In the literature more specific to health, Culyer and Wagstaff (1993), following Sen and a tradition deeply rooted in Europe, claim that health ought to be considered as a Positive Individual Right, and thus they conclude that equity demands that all equally enjoy it. The same position is echoed in recent official policy documents[179], and, for certain segments of the population, such as children, in international treaties[180]. Thus, if health is indeed a human right, the answer to our above question is straightforward: resources for

---

[178] Consider the example of a society where everybody is equal in terms of wealth, income, etc.. One may well argue that such society is inequitable because individual rewards do not reflect each person's effort. Then increasing equity, as well as efficiency, in such a society would probably require increasing inequality. Note that the distribution of all resources, assets, etc is always unequal in our societies, although to a different extent. Judging whether this inequality is also equitable or inequitable requires a value judgment based on our perception of what would be the 'fair share of the pie' that each of us should receive. In general, knowing how the pie is divided is not sufficient. An example proposed by Le Grand (1992) describes the situation where two sisters, Mary and Anne, receive an equal share of a pie, even if only Mary helped grandmother to bake it. In such a situation, giving Mary a greater piece relative to Anne to reward her work would increase equity. However, suppose that Anne could not help because she was sick. Then, sharing the pie in half may be just. In other words, we cannot judge whether or not a situation is equitable just from information concerning end-states (equal or unequal division).

[179] See, for example, WHO World Health Report 2000.

[180] For instance, the Convention on the Rights of the Child, ratified by every country with the exception of the United States and Somalia, recognizes health among the fundamental rights of all children (Art. 24).

health care should be redistributed as much as possible to the benefit of those who currently cannot enjoy this right, or cannot enjoy it to the fullest[181].

Suppose we do not consider health as a human right. Then, to assess whether the distribution of health benefits is equitable or inequitable, one could refer to the criteria utilized for distributing other important resources, such as land. Here, all different positions may be ordered along a line, at the two extremes of which we find the following two. According to the first, the distribution of assets, resources, etc...in our societies is fundamentally just, and those few instances of unfairness are a matter of unavoidable chance. Therefore, in order to be just or equitable, any externally imposed change must preserve such original evenness. For example, a tax cut needs to be proportional and preserve the relative incomes and wealth, or at a minimum not to lead to rank reversal (see Feldstein, 1976), because that relative income and wealth truly reflect each individual's relative 'merit'. According to the opposite polar position, our societies and the existing allocation of resources in them are fundamentally unjust, largely the fruit of uneven initial conditions, privilege or "undeserved" luck[182].

Applying the two opposed views to the distribution of health, health resources and services at the global level and within each country, one may argue that the existing distribution, which is everywhere strikingly in favour of rich countries and of rich people within each country (see evidence presented before) is equitable, because access to health services should be part of any society's reward system, or because the existing distribution simply reflects a high income-elasticity of demand for health services. Or, taking the opposite view, argue that the existing distribution is utterly unjust, because it reflects inequitable privileges and uneven initial conditions.

Rawl's Theory of Justice (1971) provides an insightful example of the consequences of taking the second of the two above positions. Rawls claims that it is not possible to build a theory of justice

---

[181] If we endorse the human rights perspective for health, then equity would demand that health outcomes be equalized (Culyer and Wagstaff, 1993), which, given the current distribution of burden of disease, would entail that the distribution of health inputs should be radically redistributed in favour of the poorer Nations, and in favour f the poor in each Nation.

[182] For example, considering land-ownership distribution, which is extremely unequal in several countries, according to the first perspective one would argue that such distribution is just because it reflects individual preferences, effort and marginal productivity, whilst according to the second perspective the current distribution is utterly inequitable, the result of privileges consolidated over the course of several generations. According to this second view, an equity-enhancing change would demand a radical redistribution of land in favour of the land-less and the poor, whilst according to the first one should try not to upset the existing order, and just remove impediments to the 'natural' exchanges occurring through market interaction.

based on people's current position in society, because the latter is arbitrary, and because each individual would then choose different guiding principles to maximize her/his own benefit, given her/his position. According to Rawls, if instead people could choose under a "veil of ignorance", that is to say without knowing which position they will end up occupying, they would unanimously opt for two principles, the first of which guarantees the maximum attainable level of liberties for all, and the second of which contains the Difference, or Maximin principle[183]. Such principle states that inequalities may exist if and only if they are to the benefit of the poor. In case of health this would imply that we should stop re-equilibrating health care inputs, only when further intervening would create a negative impact on the health of the poor.

Note that in our societies assessing how equitable is a certain distribution of any asset or resource ultimately requires judging how equitable market mechanisms are in allocating resources. Markets and prices perform two main tasks in our economic systems. The first one is to co-ordinate the economic behaviour of individuals. By providing signals of the relative scarcities of different commodities and assets, market prices are able to achieve allocative efficiency if certain conditions hold, as investigated for over one hundred years by economic theory[184]. However, prices in reality perform another task: they ration scarce economic resources among individuals, according to their willingness and ability to pay. Given a certain price of a good or service, two individuals, of which the first is not willing to pay that price given her preferences and the second is not able to pay that price given her wealth and income, are treated in exactly the same way by the market. They are excluded from consumption of that particular good or service[185]. The same logic applies to the interaction between countries. If -let us say- access to medicines is regulated by market mechanisms, a country who is not interested to introduce a specific drug and another who is not able to pay for it are treated in exactly the same way: they are both excluded from consumption of that particular drug.

---

[183] They would choose such second principle out of self-interest and of fear of "later" being destined to occupy the place of the most disadvantaged in society. The economist philosopher Harsanyi showed that the Difference Principle could be chosen if and only if individuals were extremely risk-averse.

[184] If certain conditions are satisfied (among these, price-taking behaviour by individuals and firms, perfect information, completeness of markets and absence of externalities and public goods), market prices determine equilibrium allocations that are Pareto optimal (first welfare theorem).

[185] In other words, any of us is able to satisfy any need or preference he/she may develop for any commodity constrained by his/her Opportunity or Choice Set, which in our societies is largely determined by our "performance" according to market criteria.

Annex 6.2 Equity in Health: The Human Rights' Perspective

As we discussed in the previous Annex 6.1, for an equity assessment concerning the distribution of any resource or asset in society it is critical to first characterize the "benchmark distribution". In general, the benchmark distribution is not perfectly egalitarian, and it can be different according to each of us. However, for everything that is considered as a human right, matters are different and more straightforward. By definition, individual human rights, unlike assets and resources, are inviolable, inalienable (one cannot buy or sell his/her right to vote, or his/her right to free speech) and must be equally enjoyed by all. Thus, perhaps the first question to ask is the following: "Shall we consider health as a human right?"

I believe that the critical distinction here is between those that consider as proper human rights only the so-called Negative individual Rights, and those that extend the realm of rights to include also Positive individual Rights. The former claim that special protection should only be guaranteed to a few core rights, such as the right to safety, to freedom of expression, to non-discrimination, etc.[186]. Those are the original human rights that were first articulated in the period of the war for Independence in the United States and of the French Revolution[187]. According to this view, governments should protect the above limited group of Negative Human Rights, and for the rest not interfere with individuals' decisions[188]. More recently, comparable arguments have been forcefully re-proposed by the noted American philosopher Nozick (1974), advocate of the so-called Minimal State Theory. According to Nozick, what matters to judge over the "fairness" of a society and over the allocation of resources resulting from social interaction, is whether or not individuals' Negative Rights have been respected, and whether the "rules of the game" are the same for everybody regardless of the initial conditions. Equity is an attribute of the process, and not of the end state.

---

[186] Such rights were considered as "natural", bestowed on each individual by God, first in the 17th century by Iusnaturalist philosophers, and then in the 18th century by the Founding Fathers of the U.S.A. and by French Enlightenment philosophers. In those times, the government mainly meant "the king", and so those theories were actually first meant to protect each person from the arbitrariness of royal power.

[187] It is interesting to see how some of the key dividing lines that would subsequently characterize the political spectrum up to the present days, described in the text, were already apparent in the interpretation given to the ideals of Liberte' and Egalite' by the different political factions that fought for power in France in the period 1789-1795.

[188] In the period of the industrial revolution, among all the Negative Rights a prominent place was assigned to property rights and to the right of free economic initiative.

On the other hand of the political spectrum are those who claim that it is what each person can actually be or do in a specific society that matters, and not so much what the government formally allows each of us to be or do. Against a formal view of freedom typical of the Negative Rights 'school', which perceives freedom mainly as the absence of interference from others, the focus here is primarily on end-states and Positive Rights. These include such rights as that to adequate nutrition, to housing, to education, to health services, to political representation, to the possibility of employment, etc. The claim is that a society that respects all Negative Rights may in fact be extremely unjust, if it *de facto* excludes part of its citizens from the possibility of flourishing as individuals because of material constraints[189]. The typical 19th century Marxist example is that of a society that protects Negative Rights, including private property and economic freedom to the fullest, but where most people have no choice, but to sell their labour under appalling conditions in order to survive. According to this second viewpoint, the answer to the question: "Equality of what?" Is: "Equality in Positive Rights and, to an extent, in end-states".

In development assessments, the shift towards putting people's Positive Rights at the centre came to be recognized with the advent of the 'Basic Needs' approach (ILO, 1976; Streeten and Burki. 1978, Streeten *et. al.*, 1981). This approach emphasizes the priority of meeting the basic needs of people, and the content of these needs is seen to include nutrition, health, shelter, water, sanitation, education and other essentials (Desai, 1990). More recently, building and elaborating over such theories and over Rawls's [190] "Theory of Justice" (1971), the noted Indian economist and philosopher Amartya Sen argued that development should be conceptualized and assessed as peoples' ability to do things that they have a reason to value, what he called the 'capability' approach' [191] (Sen, 1973, 1984, 1985, 1992). Critical to the capability argument is the notion of human freedom (Sen, 1992, 1999), and what characterizes the capability-position is again a peculiar view of freedom. Sen, distancing himself from Marx, endorses the 'intrinsic' as opposed to the 'instrumental' view of freedom (Sen, 1989). The intrinsic view values freedom for its own sake, whereas the instrumental view considers freedom to be important merely because of its significance for other achievements. However,

---

[189] The first to open this perspective on justice was Jean-Jacques Rousseau (1712-1788), in his work: "The Social Contract" (1762). Rousseau imagined that the rules governing a just society must be determined by referring to a State of Nature, where individuals still do not know which position they will occupy in society, a thesis re-affirmed in modern times by J. Rawls (1971). The first political leaders who referred to Rousseau's ideas were the Jacobin leaders, during the revolutionary government in France (1792-1794).
[190] Rawls (1971). See discussion of fourth question, hereafter.
[191] Sen (1987) argues that a person's well being is characterized not only by the state (functioning) in which a person lives or is living, but also by the existence of a range of alternative states from which the person can choose (capability). In empirical studies, it turns out that it is in fact extremely difficult to distinguish between capabilities and achieved functionings.

distancing himself also from the Negative Rights tradition, Sen specifically proposes that assessments be based on the 'positive-intrinsic' type of freedom. Thus, Sen's reply to the question: "Equality of what?" is: "Equality of capabilities across individuals in society" (Sen, 1993)[192]. For equity-based social assessments dealing with developing countries, Sen proposes a freedom-based examination of the level and distribution of relatively elementary, yet important, capabilities, of which health is a key component.

The question: "Equality of what?" and the distinction between the Negative and the Positive Rights 'schools' of thought is so important because the broader is the set of Positive Rights that society is prepared to protect, the more the government must encroach upon individual's Negative Rights in order to do so. In other words, moving towards more equality over some dimensions, in general requires moving away from it over some other dimensions. For example, increasing the range of social entitlements to fulfil Positive Rights may entail restricting individual economic freedom, or at least some individuals' faculty to spend their money in the way that they see fit. Or, imposing to individuals a level of utilization of health services compatible with achieving the right to health would entail violating consumer sovereignty[193].

Vice-versa, respecting individual preferences would entail violating equality in utilization of health services, and even more equality in health, because, given that demand income elasticity is greater than one for health services, without government intervention we would expect higher utilization and also a higher share of income spent on health services among the rich than the poor. In synthesis, there are trade-offs between different values and we, as members of society, are called to choose among them. For human rights, we should aim for equality of outcomes, whereas for other dimensions inequalities and, given the pervasiveness of trade-offs, indeed 'trade' between conflicting aims is to be accepted.

In the literature more specific to health, Culyer and Wagstaff (1993), following Sen and a tradition deeply rooted in Europe, claim that health ought to be considered as a Positive Individual Right, and thus they conclude that equity demands that all equally enjoy it. The same position is echoed in

---

[192] Sen's thesis has also been widely used by the UNDP for its 'human development' agenda See the collection of papers in UNDP (1997). Also see Anand and Sen (1994a, 1994b, 1995).

[193] Given the influence of all the other socio-economic determinants, achieving equal health expectancies across socio-economic groups would actually entail forcing a disproportionate use of services by the poorer members of society (see Annex 1).

recent official policy documents[194], and, for subsets such as child health, in international treaties[195]. Alternatively, some consider that the realm of individual Positive Rights should be limited to a minimum level of entitlements against the great risks of life, among which is illness[196]. The emphasis here in on services, rather than outcomes, and on the limits of the human rights' perspective in health to a circumscribed set of services. The extent of protection beyond the minimum should instead vary according to the level of socio-economic development, and according to the social preferences that characterize each society.

Is it possible to reconcile the two views? Culyer and Wagstaff (1993), argue that health is a Positive Right because it is one of the more important material preconditions necessary for each of us to flourish as an individual. In this respect their argument is convincing. Unlike other rights, however, being healthy or reaching old age in good health is not something that governments can equally guarantee to all. Thus, a first obvious qualification of the right to health, as Gakidou, Murray and Frenk (2000) point out, is that the maximum achievable target may in fact be to equate individual health expectancies (for instance, across socio-economic groups). Variations in healthy life spans that are simply due to chance cannot be avoided. More importantly, recent research shows that the society's socio-economic structure (its level of income inequality, for instance), and other causes external to the health sector are likely to account for the distribution of health status within a society, probably as much as, or more than the quality of medical care per se. This finding has two implications. First, that the right to health cannot be guaranteed unless other rights, such as the right to adequate nutrition, to healthy housing and healthy working conditions are also guaranteed[197]. Second, one should consider that perfect equality of health expectancies most probably couldn't be achieved, unless all people enjoy the same socio-economic condition and the same life-style.

---

[194] See, for example, WHO World Health Report 2000.

[195] For instance, the Convention on the Rights of the Child, ratified by every country with the exception of the United States and Somalia, recognizes health among the fundamental rights of all children (Art. 24).

[196] The emphasis here is on financial protection. This implies that public support should be stronger in those situations, such as in absence of family support, when people are generally unable to cope, if left alone. The British Social Security System, and in particular, the pension system, was articulated along these lines. It had from the beginning the main function of providing only a minimum safety net.

[197] A related issue is to determine along which dimension it is best to redistribute inputs (resources) and outputs. Is it better to sharpen targeting of health services, or to achieve better education for the poor, or to build safer roads in the rural areas, or greater equalization of income and wealth through taxation? Interventions channelled through sectors different from the health system may in fact achieve much more significant results in reducing health inequalities than those provided directly within the health system (see, for example, Hammer, 1998).

Thus, as we pointed out above discussing Negative vis-à-vis Positive Rights, it is a matter of trading-off the aspiration to equal health with other aspirations. But the definition of a human right excludes the possibility of trade-off. Thus, here we find a second necessary limitation to the idea of health as a human right. For health (as it is for housing or education), there in a continuum between the state of perfect health and that of death, and each society needs to draw a line between what is reasonable to consider as an individual right and what ought to be considered differently. Following this perspective, we may for instance argue that maternal and child health are indeed to be considered as human rights, so that equity demands that safe-motherhood and equality in health outcomes among children be pursued to the fullest (by implementing strongly pro-poor interventions), whilst certain differences in adult health and in utilization of health services across adults should be accepted, because eliminating them would severely limit other rights, that we consider more important [198].

Finally, we can further qualify the human rights approach to health by taking an empirically oriented perspective (see, for example, Willis, 1993, Waters, 2000). Regardless of the conclusions from a philosophical point of view as to what extent a human right to equal health should be protected, what governments can actually measure and somewhat control are access to and utilization of health services, rather than health status per se. This constraint can be rephrased by referring to the distinction between inputs, outputs and outcomes, which is so important in health. Equity objectives might be posed in terms of outputs or outcomes. However, the public sector may have control only over some inputs, but not fully on outputs and even less on outcomes. Thus, to make the right to health "operational", 'all' governments can do is to enhance access and utilization of health services by the poor, promote healthy behaviours and impose universal respect of certain health protection measures in production and consumption (at home, in the workplace, on the roads, etc.). Then, our equity focus or assessment should be more on the utilization of health services or health care, rather than on health outcomes. That is the focus we have decided to take in the text.

---

[198] Of course, the extent of health and health care that constitutes an individual positive right, or of "the minimum amount of health care entitlements" is not fixed, but evolves over time. We can draw a parallel with education: the extent of the right to education has progressively been extended over time, from simple elementary school education, to the present compulsory school curriculum. Yet, part of the higher education is still allocated "unequally" in society, according to preferences and merit.

Annex 6.3 Political and Economic Limits to Redistribution

This Annex tries to address the following questions: "Is redistribution possible at all?" "To what extent are governments capable of improving upon the allocation of resources (including health care) which would result from luck and socio-economic interaction?" One may even think that we live in a very unjust society indeed, but that any government intervention would eventually make things worse[199].

In economics, the state vs. markets research debate, one of the constant themes of contention since the mercantilism vs. *laissez faire* debate of the 18th century, has in recent years received new impetus. On the one hand, information economics has identified new areas for potential redistributive and at the same time Pareto-improving public policies [200] (in the context of health insurance, see Townsend and Harris, 1981, and Belli, 2001); on the other, new developments in micro and macroeconomic theory have further restricted the scope of government intervention in redistributing resources. For instance, Second Best theory [201] (for an application to the issue of

---

[199] Some have tried to find a way out of the state vs. markets dilemma, by advocating a greater role for voluntary, community-based channels of redistribution. In health, for example, voluntary, community-based insurance schemes have been advocated as the way forward by those dissatisfied by traditional government-led public schemes and at the same time suspicious of profit-motivated insurers. We will limit ourselves to two simple observations. First is that solidarity, unlike equity and justice, is a private virtue and that, as Sidgwick wrote (1907, p. 242): "benevolence begins where justice ends". Thus, equity-motivated (re)distribution is different and not in contrast with benevolence-based (re)distribution. Second, we should consider that what ultimately characterizes governments (centralized as well as decentralized ones) is that they dispose of a monopoly over legal coercive power. Thus, as explained in the text, if we embrace the thesis in favour of a substantial redistribution of education, health or even wealth compared to the distribution that would emerge as a result of spontaneous social interaction, we must accept public enforcement mechanisms of one sort or another to achieve it. Community-based, voluntary schemes will always find a limit in that they demand universal consensus of participants. In the language of welfare economics, they will promote only cross-subsidies that are also Pareto-improvements.

[200] Note that economic theory has traditionally focused on efficiency rather than equity issues, and has adopted a particular exception of the notion of efficiency, that of Pareto. The Pareto principle requires unanimity (at least potentially, according to the compensation principle) to accept any change as an improvement. Thus, the Pareto principle leaves us at a loss for deciding whether any change in the distribution of resources achieved by markets is to be considered as an improvement, unless the initial situation is characterised by market failures. By excluding interpersonal comparisons of utility, the Pareto principle *de facto* assigns an infinite weight to the preferences of whoever is negatively affected by any redistributive policy (for instance in case of health, those who have to cross-subsidize others' health consumption beyond the level they would spontaneously choose).

[201] The theory of Second Best has shown that under conditions of asymmetry of information the relevant efficiency concept to be used is that of "constrained Pareto-efficiency". For example, when there are two different groups of individuals characterised by unknown health risk, first best equilibria in the health insurance market are no longer sustainable, as they violate the high-risk group incentive-compatibility constraint. These constraints entail that, whenever two different contracts are offered to the two groups characterised by different health risks, they must be devised so that each group prefers the contract specific to its own risk type to the contract set for the other risk group. Second best theory has found some of the more

optimal taxation, see Diamond and Mirrlees, 1971) has showed that in general allocative efficiency and redistribution are part of the same problem. It is not legitimate to consider governments' redistributive policies, assuming that the total size of the resources available will not be affected by such policies[202]. That is to say that there can be a trade-off between maximizing the size of wealth or income, and that of redistributing it more equally (and equitably, according to many), through progressive taxation and pro poor financing/delivery of services[203]. As we discussed in Chapter 4, a comparable trade-off can arise in health, between policies aimed at maximizing gains in health outcomes, and those aimed at distributing health outcomes more equally.

To assess the distributional impact of government intervention it is useful to view the state-versus markets options as alternative mechanisms for rationing scarce resources among competing aims. We already mentioned how markets function. What about governments? Here, traditionally the two extreme positions were exemplified, on the one side, by the Public Choice Theory (Brennan and Buchanan, 1977), and, on the other side, by various modern versions of the Hegelian view of the state (after his first proponent in modern times). Public Choice theory assumes that politicians and civil servants are in general self-interested individuals who utilize their public offices to consolidate their personal power and wealth, and their decisions hardly reflect their constituencies' welfare. Even when governments pursue benevolent aims, they do so because they are instrumental to achieving politicians', or bureaucrats' real, egoistic aims[204]. On the opposite side of the spectrum, modern Hegelian views of government assume that in general the latter is able to best synthesize and sublimate the different and conflicting individual interests and preferences that characterize civil society. As expected, within a Public Choice framework, there is no room for improvement upon the results of 'spontaneous social interaction, while all the different modern political and economic positions that are rooted in Hegel's 'trust' in the role of the government reach positive conclusions as regards the latter's ability to redistribute resources in favour of the poor.

---

interesting applications in the theory of taxation (see discussion in the text)

[202] The image of public redistributive programs frequently used is that first put forth by Okun (1975) of a 'leaky buckle'. If we try to carry water using this leaky buckle from one place (rich peoples' pockets) to another (poorer people), some of the water will be lost along the way. The larger is the quantity of water we intend to transfer, the larger the amount that will be lost.

[203] Moreover the potential of taxation is severely limited in the first phases of economic development. Taxes and payroll-based contributions also introduce distortions in the private economy (see Diamond and Mirrlees, 1971, Musgrave, 1976 and Atkinson and Stiglitz, 1980).

[204] Another political view, that has a completely different origin but reaches similar negative conclusions as to the possibility of redistribution, claims that governments simply defend the interests of the underlying society's leading class, or elite group. Thus, government leaders will never champion redistribution of resources in favour of the poor, unless this is instrumental to consolidating their own group or socio-economic class's power.

In the English and American traditions, few theorists have gone as far as Hegel did to assign to the government a will and concerns independent and superior to those of the individuals that form it. However, there is a long tradition (dating back to Adam Smith) that entrusts governments with the role of correcting the so-called market failures (see discussion in Chapter 2), and also of redistributing resources and/or providing relief to the poor. Note that much of the theory that justifies governments' intervention to amend market failures implicitly also assumes that governments are benevolent welfare-maximizers. [205]

In the last two decades, new developments in political science have begun to shed some light over modern electoral mechanisms and social decision-making processes [206] that determine how real-world political and bureaucratic choices are shaped, trying to reconcile and to further articulate the above alternative views. A positive analysis of how governments function and political decisions are formed is showing that the range of actual goals pursued by public institutions have varied enormously across time and across countries, according to the underlying socio-economic and cultural structure of society, to the strength of democratic control mechanisms, and to the internal organization of governments themselves[207]. The traditional benevolent view as well as the Public Choice view, are both too abstract and simplistic.[208] In particular, the Public Choice view provides a caricature view of civil servants and politicians, and it fails to recognize that motives other than pure self-interest play an important role in determining their behaviour. In the past, several governments have been able to forcefully pursue welfare-oriented policies, redistribution of resources in favour of the poor, and defence of human rights, in a way not reconcilable with a solely cynical and self-interested view of public decision makers.

---

[205] Such an assumption is still at the basis of several contributions in economic theory. For instance, Contract Theory and Theory of Regulation microeconomic models for the most part assume that the public sector, although subject to the same information constraints as private agents, maximizes welfare (see for example, Laffont and Tirole, 1993. See however, theory of Regulatory Capture in the same textbook).

[206] In economics, at the basis of the recent developments we find the median voter theory, which explains the move "towards the centre" observed in all uninominal electoral systems, and Arrow's seminal work on the so-called Impossibility Theorem (Arrow, K., Social Choice and Individual Values, 1951). Arrow (1951) was the first to explicitly investigate the issue of aggregation of individual preferences in social choices.

[207] Following is a partial list of the range of criteria that governments can follow, and have indeed followed, in the allocation of resources across individuals. 1) Same as markets, willingness and ability to pay; 2) political leverage, 3) military force; 4) first-come-first served; 5) same quantity for everybody. (equality of consumption across households was more or less achieved in rationing bread and sugar in Europe during world war II); 5) some measure of economic need; 6) age and gender; 7) capacity to benefit 8) capacity to bear queues; 9) chance; 10) a combination of the above.

[208] Results derived in economics under the "government-welfare-maximizer" assumption can be misleading, because they ignore the concrete political, economic, and administrative constraints.

The important point to understand, though, is that dispensing with markets does not automatically guarantee a different and more equitable allocation mechanism. Markets may not lead to equitable or efficient outcomes, but still:

1) Their limitation/elimination poses enormous problems of finding and sustaining alternative resource allocation mechanisms.

2) It is not at all guaranteed that public intervention will achieve more equitable results. The evidence presented in this Chapter on the distribution of health benefits from publicly funded services in developing countries is a clear illustration of this point.

Chapter 7: Ten years of health reforms in former socialist economies

7.1 Introduction

This Chapter presents a study and an assessment of health reforms in former socialist countries of Central and Eastern Europe (CEE) and the Former Soviet Union (FSU or CIS), hereafter former socialist economies (FSE), thus setting the stage for the more detailed analysis of the Georgian case in Chapter 8. In the first part, it provides a synthetic picture of the health systems inherited from the communist regimes, and it presents the macro variations in health status, public health expenditure, private funding, etc...that occurred over the transition years. This first part draws together evidence from several studies and it is intended to provide an information basis for the subsequent analysis of the health reforms.

7.2 Health Threats and Health Status

Under the communist regimes imposed everywhere after 1945 in Central-Eastern Europe and in the Soviet Union health care was declared a public responsibility. By the mid '70s the goal of universal access to care with broad coverage in terms of services was largely achieved by most socialist countries in the region. Their health systems also provided strong public health measures, such as compulsory childhood immunization, which helped them to achieve better health outcomes than other countries with similar levels of income[209]. By the mid-1960s life expectancy at birth in socialist countries was comparable to that in the West.

However, beginning in the second half of the 70s aggregate health status indicators showed no further improvement and adult health indicators in fact started to worsen. Such stagnation and deterioration continued in the '80s, and in some of the countries in the region accelerated over the transition years, with the more well-known and dramatic example involving male life expectancy in the Russian Federation (see Case study 7.1). Largely because of increased deaths from cardio-

---

[209] Poland for example, during the period following WWII achieved the best results in term of reduction of infant mortality rates (IMRs went down to 40 per 1000 in the '60s from 110 per 1000 in the '50s (Zatonsky, 1996, "Evolution of health in Poland since 1988").

vascular disease [210] and external causes among middle-aged men (mainly deadly injuries, including murder and suicide, and alcohol self-poisoning), male life expectancy at birth in Russia declined 7.4 years (from 64.9 to 57.5years) between 1987 and 1994—an extraordinary change[211].

Case study 7.1: Life expectancy in the Russian Federation

Recent evidence shows that, between 1985 and 1987, Russian life expectancy increased from just below 68 years to 70 years (McKee and Leon, 1997). Then it reversed, plunging to 64 by 1994 (and to 57.5 for men), before rising slightly in 1995. The improvements after 1984 reflected reduced deaths from many causes, notably accidents and violence, alcohol-related causes, heart disease, and pneumonia. The improvements were concentrated among young and middle-aged adults, especially those aged 40-44 years. The subsequent plunge involved a reversal of all these trends. The worst years were 1992-1994, when men lost 4.5 years of life expectancy. An official Russian report in 1997 by an expert panel of demographers addressed this unprecedented increase in mortality in the Russian Federation (Pres. Comm. on Problems of Women, Family, and Demography, 1997, *The Current Mortality Situation of the Population of Russia*). It presented the following bleak picture of health status in the Federation: Russian men live 15-17 years less, and women 7-10 years less than their Western counterparts. 82% of the recorded increase in mortality reflects a real increase in age-specific death rates, while only 18% can be attributed to age-structure effects. The subsequent decline in the number of deaths between 1995 and 1997 was largely due to effects of age structure and did not represent a real decline in death rates. In 1997 life expectancy in the Russian Federation was approximately equal to 66.5 years.

The Russian Federation case, although particularly striking, is paradigmatic of a trend shared by all the countries in the FSU (which became Community of Independent States, or CIS, after the collapse of the union in 1991). In Georgia, for example, maternal mortality rates increased by approximately 45% between 1990 and 1997; deaths caused by cardiovascular disease increased by 35% and the overall age-adjusted mortality rate increased by 18%[212].

By contrast, over the last decade some of the Central and Eastern European (CEE) countries were able to reverse the previous trends and to achieve significant improvements in life expectancy. The improvements in life expectancy were particularly strong in Slovenia, Czech Republic, Slovak

---

[210] Standard death rates for circulatory disease among Russian men rose from 658 per 100,000 in 1991 to 837 per 100,000 in 1995. Among men, the standard death rate from CVD in Russia is higher than the death rate in the U.S. for all causes combined (Eberstadt, 1999).

[211] Between 1987 and 1994, Russia experienced a 250% increase in adult male (age cohort 25-44) age-specific-death rates (ASDRs), and a 200% increase in adolescent and adult female (age cohort 15-44) ASDRs (Shkolnikov, 1997).

Republic and Poland. In Poland, in the first two years of the transition up to 1991, mortality rates continued to worsen, especially due to externally caused deaths that rose by 25% from '89 to '91. Then, there was a sharp as well as unexpected improvement, and from 1991 to 1994 the overall mortality rate decreased from 1,062 to 971 per 100,000 for men and from 550 to 519 per 100,000 for women[213]. The starkest phenomenon was the decrease in deaths from cardiovascular disease, which fell by 25%, in the 20-44 age group, and by 15% in the 45-64 age group. Over the same years comparable positive trends were observed in the Czech and the Slovak Republics[214].

In other countries of CEE, however, over the decade health indicators continued to deteriorate, although the setback in health indictors was not as sharp as in the FSU. A negative trend was observed not only in all the war-stricken states of the Balkan region, but also in fast growing Hungary (see table 7.1).

Throughout the FSE the leading cause of death has been from cardio-vascular disease[215]. In the worst cases, such as in Hungary, age standardized death rates (SDRs) due to circulatory disease and ischemic heart disease among men are nearly double those of the United States. Cancer— particularly lung cancer—and injuries are also unusually common among adult males. Ischemic heart disease, cerebrovascular disease and lung cancer together explain a third of the total number of years lost in the European region of the FSU.[216] Another leading cause of death has been from external causes, mainly injuries and self-poisoning. By contrast, in the Central Asian region communicable diseases (mainly respiratory infections and TB), perinatal and maternal causes are still responsible for approximately half of the total burden of disease.[217]

---

[212] See Georgia Public Expenditure review, World Bank, 2002.

[213] Zatonsky, 1996, "Evolution of health in Poland since 1988".

[214] Albrecht, V., Drbal, C., Zdravotnictví v ČR: krize nebo úspěch? Zdravotnictví v České republice, 3/1998, s.10-13

[215] According to Murray and Lopez (1997) cardiovascular disease is responsible for 26.2 of the total number of DALYs lost in former socialist economies.

[216] Bobadilla and Costello, 1997.

[217] Bobadilla and Murray, 1997, estimate on the basis of 1990 data that communicable disease and maternal and child deaths constituted 53% of total burden of disease in Central Asia. Such share does not seem to have significantly varied over the transition.

Table 7.1: Aggregate Health Status Indicators for a Selected Group of FSE[218]

	Male life expectancy			Female life expectancy			Infant mortality rate		
	1990	1993	1997	1990	1993	1997	1990	1993	1997
Albania	69	69	70	75	74	75	31	33	27
Azerbaijan	67	65	67	75	74	75	23	28	23
Croatia	69	68	68	76	77	77	11	10	9
Czech Rep.	68	69	71	76	76	78	11	9	6
Estonia	64	62	64	75	74	76	12	16	10
Georgia	**69**		**69**	**76**		**77**	**16**	**22**	**22**
Hungary	65	65	66	74	74	75	15	13	10
Kazakhstan	64	62	60	73	72	71	26	28	24
Kyrgyz R.	64	63	63	73	71	71	30	33	28
Latvia	64	62	64	75	74	75	14	16	15
Moldova	65	64	63	72	71	71	19	22	20
Poland	67	67	69	76	76	77	16	16	11
Romania	67	66	65	73	73	74	27	23	22
Russian Fed.	64	59	60	74	72	72	17	20	18
Slovak Rep.	67	68	69	76	77	77	13	11	9
Slovenia	69	70	71	77	77	78	8	7	5

Sources*: Jakab et al., 2001; Kornai, 2000; Shkolnikov, 1997; World Development Indicators, 2002*

Note that women continue to fare better than men in terms of absolute levels of mortality and morbidity: in 1990, women accounted for 52% of the population, but only for 43% of the estimated DALYs[219] lost. However, of particular concern is the sharp increase in female lung cancer mortality, which demonstrates the very real threat to women's health posed by smoking –a threat that is expected to worsen as the tobacco industry continues its efforts to expand its market through targeting of women and young people. In the '80s and '90s death rates from cervical and breast cancer also continued to rise, and the dearth of screening services and prevention programs for cervical and breast cancer around the region is a serious matter of concern. As concerns breast cancer, for example, WHO estimates that 30-40 percent of cases in Estonia, and almost 2/3 in Poland are still detected at more advanced, and more difficult to treat, stages. Other countries, for which

---

[218] It has not been possible to check the reliability of the primary source.
[219] Disability Adjusted Life Years, first used by the World Bank in 1992 (WDR, 1993), is a measure that takes into account not only mortality but also morbidity.

official figures are not available, are likely to perform even worse at early cancer detection. Furthermore, the state of reproductive health services is appalling, with a few positive exceptions among CEE countries such as the Czech Republic and Slovenia[220]. In most countries of the region access to modern contraceptive products and accurate information regarding family planning alternatives is privy to a small minority of the population[221] and abortion remains the dominant method of fertility control. Contraceptive prevalence rates in women aged 15-49 are 34% in the Russian Federation and 48% in Romania. By comparison, in Thailand, a country at a similar level of economic development, it is equal to 74% (World Bank, 2002). It is estimated that on average a woman living in the Former Soviet Union undergoes over 3 abortions during her fertile life (for instance, in Georgia the estimate is equal to 3,7[222]). The rate of abortions to live births is more than one for the majority of FSE. The largest ratio is in Azerbaijan, with official data showing 2,199 abortions for 1,000 life births (1995 data)[223].

As for infant mortality, during the transition years the trend has been negative in several countries of the FSU, whilst has been stagnant or positive in CEE countries. In some of the FSU countries the relatively good performance according to official figures (see table 1, based on vital registration data) is not confirmed by survey-based evidence. In the Caucasus and in the Central Asia Republics, for instance, over the decade the cases of undefined diarrhoea diseases, meningitis, infectious mononucleosis, rubella and influenza sharply increased. Thus, one would expect that also infant mortality rates deteriorated and not that they improved, as the official figures show in some of the countries. On average in the FSU infant mortality is equal to approximately 20 deaths per 1,000 children, a value almost four times higher than in the industrialized countries. Moreover, the aggregate country data mask wide variations across socio-economic groups and geographic areas

---

[220] As regards maternal mortality, over the transition years it does not seem to have grown significantly, and the best performing countries have managed to reduce it. Maternal mortality averages 41 deaths per 100,000 live births in the FSU countries, and 21 per 100,000 in Central and East Europe. The range of variation, however, is quite broad; it extends from 17 per 100,000 in Lithuania, to 50, 65 and 70 per 100,000 in Russia, Kyrgyz Republic and Kazakhstan respectively, up to 96 per 100,000 in Tajikistan (1995 data). Within Central and East Europe, Romania continues to perform worse than any other country, with 41 deaths per 100,000, while estimates for Bulgaria, Hungary and the Czech Republic are respectively 15, 12, and 9 respectively. Finally, a rising concern is posed by female death rates from cardiovascular disease that remain significantly higher than rates in the European Union (in Central and East Europe 86.3 per 100,000 and in the FSU 98 per 100,000, as opposed to 33.1 per 100,000 in the EU).

[221] Major causes preventing a more widespread use of contraceptives are: 1. recurrent interruptions in the flow of supplies; 2. physicians' resistance to encouraging the use of modern contraceptive methods; 3. persistence of deep-rooted cultural and social taboos against the use of modern contraceptives.

[222] UNICEF, 1993, Georgian Women's Health Profile.

[223] Even those FSU countries where rates of abortion were lower have witnessed a sharp increase in the number of abortions over the last decade. In Armenia, the rate of reported abortions climbed from 316 in 1990 to 627

within countries. For instance, in the Kyrgyz Republic according to a Demographic and Health Survey conducted in 1997 (World Bank 2002) infant mortality was equal to 46 per 1,000 in the richest wealth quintile, and 83 per 1,000 in the poorest quintile.

In addition, within the region dormant threats from communicable diseases that affect mainly the adult population re-emerged. In the FSU, and particularly in the Central Asian republics, in Moldova and in the Caucasus, the incidence rates of tuberculosis, diphtheria, and sexually transmitted diseases rose to many times their prior levels[224]. In Russia, the TB case notification rate has doubled between 1991 and 1997, reaching 123/100,000 in 1999 (World Bank, 2002)[225]. Particularly worrying is the situation among Russian prisoners[226]. HIV infection has also increased sharply over the last few years (420,000 reported cases in 1999, according to UNAIDS, 2000), especially among injecting drug users. Syphilis incidence is estimated to have increased up to 50-fold in the last 10 years[227].

7.2.2 Epidemiological Studies on the Causes of the Deterioration in Health Status

According to the existing epidemiological studies on mortality and morbidity the growth in premature death from non-communicable diseases, observed in most FSE countries, is largely due to lifestyle factors, such as alcohol and tobacco consumption, diet, stress and lack of exercise. A global study[228] of risk factors associated with premature loss of years of healthy life (measured by DALYs) found that in 1990 in the FSE tobacco accounted for 12.5 percent of all DALYs lost, and alcohol accounted for a further 8.3 percent. The third risk factor in order of importance was hypertension (5.9 percent), which is also aggravated by the first two. According to this study the first two risk factors, tobacco and alcohol consumption, together had a larger impact than the eight next most important risk factors combined. These results are confirmed by a 1997 WHO study, according to which middle-aged men in FSE are twice as likely than those in Western Europe to die from

---

per 1,000 live births in 1995.

[224] The figures reported in the text are likely to be underestimates of the true incidence rates, as several cases of TB and STD are not reported.

[225] The TB mortality rate in 1996 was 17/100,000 (in the US is 0.61/100,000). So, a Russian is on average 28 times more likely to die of tuberculosis than an American.

[226] TB is concentrated among marginalize groups. It is the major killer among Russian detainees, most of which await due process for several years. 700 per 100,000 citizens are currently in jail in Russia, five times more than in Europe (Farmer, P., 1999).

[227] Russian Federation AIDS & TB Project Concept documents (1999).

[228] Murray, C. and A. Lopez, 1997.

tobacco-related causes. In 1995 tobacco caused an estimated 41 percent of all deaths among men aged 35-69 years[229].

Several studies suggest that the second risk factor in order of importance is alcohol consumption. For the Russian Federation, Shkolnikov and Nemtsov (1997), and Trems (1997) present the results of a number of studies on the effects of alcohol consumption on the burden of disease. Shkolnikov and Nemtsov (1997) attribute the oscillating patterns of accidental poisoning[230] and accidents in the second half of the '80s primarily to sharp fluctuations in the availability and consumption of alcohol (both licit and illicit), during and after Mikhail Gorbachev's 1985-87 campaign to reduce its availability (1985-1987). By contrast, Treml (1997) argues that the effects of Gorbachev's anti-alcohol campaign have been largely overstated. He suggests that the campaign encouraged illegal distillation and therefore resulted in even worse health outcomes, due to the poor quality of the vodka illegally produced. This could explain the fact that the two years 1985-1987 marked a significant health improvement for the well educated segments of the population, but worsening health outcomes for the lower socio-economic groups (Shkolnikov, 1997). The lack of reliable data for this period makes it impossible to reach definitive conclusions.

The level of alcohol consumption has continued to increase over the '90s, and it has now reached astonishing levels[231] in several FSE countries. Alcoholism is likely to pose an immense burden on society, and the number of preventable deaths directly ascribed to alcohol consumption to represent just the top of the iceberg of such burden.

Poor nutrition seems to be another important, yet still neglected, factor in explaining the deterioration in health status over the last few years. Under-nutrition and micronutrient malnutrition [232] particularly affect the poorer and more vulnerable segments of the population, and the poorer countries, such as those in the Central Asia region (see World Bank, 2000). Moreover, in FSE diets have traditionally exhibited iron and iodine deficiencies[233] and by the problem of non-availability and

---

[229] WHO, Regional Office for Europe, Fact sheet No 156, May 1997

[230] Most cases of self-poisoning occur as a result of excessive vodka consumption in a short period of time.

[231] According to a 1993 national household survey, over 80 percent of Russian men were drinkers and their alcohol consumption averaged nearly 600 grams per day, roughly equivalent to five bottles of vodka per week, every week (Eberstadt, 1999). For the Russian population as a whole, average alcohol intake is approximately 400 grams per day.

[232] Under-nutrition is defined as the inadequacy of food intake to meet a person's energy and nutrient needs for healthy functioning. Micronutrient malnutrition is defined as the insufficiency of essential vitamins and minerals in what may be an otherwise healthy diet.

[233] Those could easily be addressed by fortifying salt with iodine and wheat flour with iron, as is done in many other countries. Goiter, a disease stemming from iodine deficiency disorder, prevalence rates of up to 25% have been reported in Bosnia, up to 40% in parts of Russia, and up to 75-80% in parts of the Caucasus and Central Asia regions. Iodine deficiency causes impairment of cognitive capabilities, with severe long-term

non-affordability of vitamin-rich eatables in the countries characterized by a colder climate. In a few countries of CEE and among the wealthier segments of the population the opposite problem of over-nutrition has also worsened over the transition years[234].

7.3 Quality and Efficiency of the Delivery System

This section focuses on the delivery system and offers a synthetic picture of the situation inherited from the communist regimes. Before the transition to a market economy, FSE's health systems shared comparable roots, and a few key characteristics. Resources were used quite inefficiently. First, the public health delivery systems were extremely rich in facilities and labour (see Tables 7.2 and 7.3 below, and Jacab and Preker, 2002), at a level comparable to that of the richest countries of Western Europe and North America.

---

effects.

[234] Over-nutrition is defined as an excessive intake of calories relative to energy requirements and as an unbalanced intake of other nutrients (generally excess fat and insufficient fibre intake).

Table 7.2: Hospital Beds per 1,000 People in Selected Central and East European, Former Soviet Union, OECD and Newly Industrialized Nations (1990)

Central and Eastern Europe	Beds per thousand	Former Soviet Union	Beds per thousand	OECD countries and newly industrialized nations	Beds per thousand
Albania	4.03	Armenia	8.6	Austria	9.90
Bulgaria	10.10	Azerbaijan	9.9	Denmark	4.70
Croatia	7.38	Belarus	13.16	France	8.70
Czech Rep.	8.10	Estonia	11.60	Germany	9.60
Hungary	10.14	**Georgia**	**9.73**	Greece	5.00
FYR Macedonia	6.34	Kazakhstan	13.94	Ireland	4.90
Poland	6.60	Kyrgyz Rep.	11.96	Italy	6.50
Romania	8.92	Latvia	14.04	Netherlands	12.0
Slovak Rep.[235]	7.45	Lithuania	12.41	Sweden	7.60
Slovenia	6.05	Moldova	13.15	UK	6.59
Yugoslavia, Fr.[236]	5.92	Russian Fed.	13.02		
Average	7.35	Ukraine	13.49	Canada	4.20
		Uzbekistan	12.43	Japan	16.20
				USA	4.50
		Average	12.11	Average high income countries	7.7
				Korea	4.60
				Turkey	2.50

Source: *World Bank, World Development Indicators, various years*

---

[235] In fact in 1990 the Czech Rep. and the Slovak Rep. were part of the same country, Czechoslovakia. Nonetheless disaggregated data is available for the two parts of Czechoslovakia.
[236] In 1990 Croatia and Slovenia were still part of Yugoslavia. Nonetheless disaggregated data is available for the two, and it is presented separately from Serbia and Montenegro.

Table 7.3: Physicians per 1,000 People in Selected Central and East European, Former Soviet Union, OECD and Newly Industrialized Nations, 1990

Central and East Europe	Number of physicians per thousand	Former Soviet Union	Number of physicians per thousand	OECD countries and newly industrialized nations	Number of physicians per thousand
Albania	1.3	Armenia	3.3	Austria	3.0
Bulgaria	3.5	Azerbaijan	3.6	Denmark	3.4
Croatia	2.2	Belarus	4.2	France	2.9
Czech Rep. [237]	3.0	Estonia	3.2	Germany	3.4
Hungary	3.2	**Georgia**	**4.5**	Greece	3.9
FYR Macedonia	2.3	Kazakhstan	3.5	Ireland	2.1
Poland	2.3	Kyrgyz Rep.	2.9	Italy	5.9
Romania	1.7	Latvia	3.4	Netherlands	3.1
Slovak Rep.	3.5	Lithuania	3.9	Sweden	3.1
Slovenia	2.3	Moldova	3.4	UK	1.7
Yugoslavia, Fr. [238]	2.0	Russian Fed.	4.1	Portugal	3.2
		Ukraine	3.5	Canada	2.1
Average	2.5	Uzbekistan	3.1	Japan	1.8
				USA	2.6
		Average	3.6	Average high-income countries	2.9
				Korea	1.1
				Turkey	1.1

Source: *World Development Indicators, most recent year available*

Tables 7.2 and 7.3 compare countries of Central and East Europe and of the Former Soviet Union to those of Western Europe and other middle-income market economies (Korea and Turkey), in relation to respectively the ratio of hospital beds and physicians to populations. The comparisons are illuminating. All FSE were characterized by a very extensive delivery system. In Central and Eastern Europe figures were comparable with those of the West, while in CIS countries, the delivery system was really elephantine, especially in relation to their low per capita income (on average, approximately 1/10th of that of established market economies). The few crude efficiency indicators available indicated extremely low levels of efficiency. For instance, average length of stay (LOS) in

---

[237] See note previous table.

hospitals was much longer than in the West, reflecting a misuse of acute hospital for chronic or long-term care.

Second, inpatient care was taking a disproportionate share of the public budget (on average, approximately 60-65% of the total). In FSE the health delivery systems were based on an efficient network of mostly single-specialty hospitals, according to the *Siemaszko* model of the Soviet Union. A vertical and segmented approach to disease management characterized such model, which might have been originally appropriate to address communicable diseases but that certainly became outdated and unsustainable as the disease burden shifted mostly towards chronic diseases in the '60s. Moreover, there was no system of peer-review or other means for internal and/or external quality control.

Third, the system was characterized by a very convoluted governing structure, where different health providers were accountable to different ministries[239] and different levels of government, federal, regional, district and city level.

Fourth, the FSE's health systems completely lacked incentives for individual workers, as well as for institutions, to improve the level and the quality of services, or to improve efficiency. Given the criteria for the allocation of funds across facilities, which was based on existing capacity, historical expenditure and reported costs[240], the management of health facilities had few incentives to develop strategic planning functions, to introduce innovation or to adapt to environmental changes. Salaries were low and seniority-based, career progress was unrelated to performance but to political affiliation.

Thus, taking the image of an aircraft to describe the health systems inherited from communist regimes, one may say that, although to a different extent, in all FSE countries the plane was too big compared to need, was highly ineffective in its engines, unsafe and uncomfortable, and was often flying in the wrong direction. When the fuel -the money that the government had available to sustain this cumbersome delivery system- signalled red at the end of the eighties, the need for health reforms able to tackle the issues of excess capacity and inefficiency became compelling.

---

[238] See note previous table.

[239] According to rough estimates, in several FSE total medical resources at the disposal of all other ministries were almost equivalent to those under the jurisdiction of the MOH (Langenbrunner, Wiley, 1999). Co-ordination among ministries was extremely poor.

[240] No use was made of epidemiological or utilization indicators (with the exception of occupancy rates).

7.3 Health Financing

Four questions about the evolution of health financing over the last decade of particular interest are:

1. How economic growth varied in the different countries and what impact it had on total government health care spending;

2. How the proportion of the total financed publicly (through taxes and social security contributions), and privately (mainly through OOP payments) varied over the decade.

3. How the incidence of OOP payments varied across different socio-economic groups and how severe was the financial barrier to access health services.

For most FSE, reliable data are now available for GDP growth, and total government expenditure in health care, but not for total private OOP payments, particularly the informal component.

Looking first at government health expenditure and its relationship with the evolution of income and fiscal revenues, one can see that socialist countries started from similar levels of spending in proportion to their gross domestic product (GDP), and then tended to diverge over time. In fact, it is possible to distinguish between two periods and two groups of countries. In the first years of the transition, as Table 7.5 shows, almost all FSE experienced severe economic recessions. The weakness of the new democratic and market economy institutions[241] compounded the structural economic difficulties due to the interruption of the trade and the subsidies within the Soviet Union and with neighbouring socialist countries. The governments' revenue collection capacity was blighted, and in some FSU countries, such as Georgia, the extent of the collapse in income and in fiscal revenues was enormous, in the order of 50, 60 per cent.

Beginning in 1993-94, most countries in the region started to experience some economic recovery. Then, a divergent pattern of evolution between most CEE and FSU countries emerged. While the majority of CEE countries were able to move towards a path of stable economic growth, the Community of Independent States that formed after the collapse of the Soviet Union (with the partial exception of the Baltic States) in 1997 became trapped in another period of economic turbulence[242],

---

[241] Under the socialist regimes, the problem of securing enough funds to finance public goods and welfare state services was framed under a totally different institutional setting. For example, personal taxation was largely irrelevant as the state put most of the burden of financing social services directly on enterprises.

[242] The crisis began in 1998, when the East Asia financial crisis spread to Russia. In August the government announced a default on its public debt obligations. This crisis initiated a new period of economic recession in all countries of the FSU.

which terminated only in 2000/01. The main reasons mentioned in the literature to explain the different economic evolution followed by the two groups of countries during the transition years concern the extent of corruption and the way the privatisation process was carried out, the absence of rule of law in the FSU countries, and the geographical distance from European Union markets.

Table 7.4: GDP Growth for a Selected Group of FSE (annual real %)

	1990	1991	1992	1993	1994	1995	1996	1997	1998
Albania		-27.7	-7.2	9.6	9.4	8.9	9.1	-7.0	8.0
Armenia	.	-11.7	-41.8	-8.8	5.4	6.9	5.9	3.3	7.2
Azerbaijan	-11.7	-0.7	-35.2	-11.8	-24.3	-11.8	1.3	5.8	10.0
Belarus	-2.2	-1.2	-9.6	-7.6	-12.6	-10.4	2.8	10.4	8.3
Bulgaria	-9.1	-8.4	-7.3	-1.5	1.8	2.9	-10.1	-7.0	3.5
Croatia	.		-11.7	-8.0	5.9	4.8	5.9	6.8	2.5
Czech Republic	-1.2	-11.6	-4.5	0.1	2.2	5.9	3.8	0.3	-2.3
Estonia	-7.1	-8.0	-21.2	-8.5	-1.8	4.3	3.9	10.6	4.0
Georgia	**-14.8**	**-20.1**	**-40.3**	**-39.4**	**-11.4**	**2.4**	**10.5**	**11.0**	**2.9**
Hungary	-3.5	-11.9	-3.1	-0.6	2.9	1.5	1.3	4.6	5.1
Kazakhstan	-4.6	-6.8	-13.0	-9.6	-12.6	-8.2	0.5	1.7	-1.9
Kyrgyz Republic	6.9	-9.1	-15.8	-16.3	-20.1	-5.4	7.1	9.9	3.6
Latvia	-1.3	-10.4	-34.9	-14.9	0.6	-0.8	3.3	8.6	3.6
Lithuania	9.5	-5.7	-21.3	-16.2	-9.8	3.3	4.7	7.3	5.1
Moldova		-16.0	-29.1	-1.2	-31.2	-1.4	-7.8	1.3	-8.6
Poland	-9.9	-5.5	3.1	4.3	5.1	7.0	6.0	6.8	4.8
Romania	-5.7	-12.9	-8.8	1.5	3.9	7.1	3.9	-6.6	-7.5
Russian Federation	-3.0	-5.0	-14.5	-8.7	-12.6	-4.1	-3.4	0.9	-4.6
Slovak Republic	-2.7	-14.6	-6.7	-3.7	4.9	6.9	6.6	6.5	4.4
Slovenia	.			2.8	5.3	4.1	3.5	4.6	3.9
Tajikistan	-2.4	-8.7	-29.0	-11.3	-19.0	-11.8	-4.4	2.3	8.2
Ukraine	-6.4	-8.7	-9.9	-14.2	-22.9	-12.2	-10.0	-6.0	-1.7

Source: *World Development Indicators, 2000.*

Table 7.5: GDP Comparison at the Beginning and at the End of the Decade, and Average Annual % GDP Growth for a Selected Group of FSE ($ millions[243])

	GDP 1990	GDP 2000	Average annual real % growth
Georgia	**12,171**	**3,029**	**-13.0**
Moldova	10,567	1,286	-9.7
Ukraine	91,327	31,791	-9.3
Azerbaijan	9,837	5,267	-6.3
Russian Fed.	579,068	251,106	-4.8
Kazakhstan	40,304	18,230	-4.1
Kyrgyz Rep.	2,951	1,304	-4.1
Latvia	12,490	7,150	-3.4
Lithuania	13,254	11,314	-3.1
Bulgaria	20,726	11,995	-2.1
Armenia	4,124	1,914	-1.9
Belarus	35,203	29,950	-1.6
Romania	38,299	36,719	-0.7
Estonia	6,760	4,969	-0.5
Croatia	18,156	19,031	0.6
Czech Rep.	34,880	50,777	0.9
Hungary	33,056	45,633	1.5
Slovak Rep.	15,485	19,121	2.1
Slovenia	12,673	18,129	2.7
Albania	2,102	3,752	3.3
Poland	58,976	157,739	4.6

Source: World Development Indicators, 2002

As a result of the economic difficulties, initially all FSE had to cut real public spending for health, and they did so roughly in proportion with the GDP decline, as shown in Table 7.6. The exceptions are a few countries that experienced war during this period (for example, Georgia), where public spending on health essentially collapsed along with the government's revenue base[244], and, on the

---

[243] GDP is computed at purchaser prices. Conversion in U.S. dollar is done by applying the average official exchange rate reported by the International Monetary Fund for the year shown.

[244] In these countries even basic services, such as immunization services and prenatal care, which had reached standards comparable to those in Western Europe in the 60 through the 80s, could not be sustained. According to World Bank, 1999, in 1998 in Georgia only 63% of children were immunized against measles, and none in 1994-95, because of lack of vaccines. In several of the countries of the FSU and particularly in the Central Asia

opposite side of the spectrum, a few countries in Central Europe (such as the Czech Republic) where health spending could be protected even in the early transition period. In general, governments reduced planned spending, but also ran deficits, and payment arrears began accumulating in the system.

Table 7.6: Health Expenditure, Public (% of GDP) for a Selected Group of FSE

Country Name	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Albania	3.3	4.4	3.0	2.7	2.5	2.5	2.6	2.7	2.5	2.0
Armenia	2.8	4.1	4.9	3.6	1.5	2.7	3.4	3.2	3.4	4.0
Azerbaijan	2.6	3.1	2.3	3.0	1.6	1.1	1.2	1.0	1.0	1.0
Belarus	2.6	3.1	3.4	4.7	5.2	5.3	5.1	4.9	4.6	
Bulgaria	4.1	4.2	5.3	4.8	4.0	3.6	3.1	3.4	3.5	3.9
Croatia	9.7	10.8	9.9	9.3	8.5	8.4	9.3	8.7	9.0	9.5
Czech Rep.	5.2	5.1	5.1	6.9	7.1	6.9	6.6	6.5	6.4	6.6
Estonia	2.1	7.5	5.3	5.2	5.8	5.8	5.2	4.6	5.1	5.1
Georgia	**3.0**	**4.1**	**2.4**	**0.4**	**0.3**	**0.6**	**0.9**	**0.7**	**0.7**	**0.8**
Hungary	5.7	6.6	6.6	4.8	5.2	4.9	4.7	4.5	4.7	5.2
Kazakhstan	3.2	4.3	2.1	2.3	2.2	3.0	2.5	2.3	2.4	2.7
Kyrgyz Rep.	4.2	3.6	3.2	2.6	3.5	3.5	3.1	2.9	2.7	2.2
Latvia	2.5	2.6	2.8	4.1	3.6	4.4	4.0	3.5	3.9	4.0
Lithuania	3.1	5.2	3.8	3.9	4.8	5.1	6.1	5.0	6.2	4.8
Moldova	4.4	4.0	3.8	4.4	6.2	5.8	6.2	4.8	3.6	2.9
Poland	4.7	4.8	4.9	4.5	4.1	4.2	4.6	4.8	4.2	4.7
Romania	2.8	3.3	3.6	3.0	3.3	3.6	2.9	3.1	3.4	3.8
Russian Fed.	2.5	2.4	2.5	3.3	5.2	4.4	4.1	4.5	4.5	4.6
Slovak Rep.	5.0	4.6	4.6	5.8	6.3	6.1	5.8	5.8	5.7	5.7
Slovenia		5.2	7.3	7.7	7.7	7.1	6.6	6.6	6.8	6.7
Ukraine	3.0	3.3	3.5	4.1	5.4	4.9	3.9	4.1	3.5	2.9

Source: *World Development Indicators, 2000-2001*

In order to respond to the general budget shortfalls and to inject additional resources in the health system, several FSE created or increased the rate of a payroll tax earmarked to fund health services. The tax was first introduced in Hungary (1991), and Czech Republic (1992), and then several other

---

republics the administration of vaccines was completely left to International NGOs, and immunizations services left out approximately one-third of the total number of children (30% in Kyrgyz Republic and 20 in Uzbekistan, according to World Bank, 2000), mostly from the poorer backgrounds and the more remote areas.

countries followed, including Croatia, Slovenia and Macedonia, the Baltic States, Georgia (1995), Russia and Kazakhstan, and finally, Poland (1999). These countries also created sickness or insurance funds, separated from the Ministry of Health, to administer the revenue from the payroll tax. The burden of the tax was shared by employers and employees, while the central or local governments were supposed to contribute either for the non-contributing population[245], as in Russia or Hungary, or for specific services, as in Georgia. In most of the Former Soviet Union the tax was kept at a relatively smaller rate[246] (Russian Fed., 3.6%; Kazakhstan, 3.25%, Georgia, 4%; Kyrgyz Republic, 2.5%), whilst in most CEE countries and the Baltic states the payroll tax rate was set at a much higher level (Croatia, 18%; Hungary, 18%; Czech Rep., 13.5%; Slovak Rep., 13.75%; Estonia, 13%, Poland, 7.75%, Romania, 14%). In the former group of countries, the revenue of the payroll tax was not able to turn around the disastrous public health financing situation. Due to low tax rate, low and shrinking levels of formal employment in the economy, poor collection, and misappropriation, the revenue that providers could receive from the sickness funds remained negligible. For example, in the Russian Federation, in 1993 the government mandated that local budgets contribute to the mandatory insurance fund on behalf of the non-contributing population; two years later, in 1995, already 71 percent of all public resources for health originated from local budgets (of the rest, 10 percent came from the federal budget and only 19 percent from payroll taxes collected at the central level). The fiscal capacity and willingness to contribute varied enormously across regions and no mechanism to cross-subsidize the poorer regions was put in place.

Table 7.7 presents the total private and public health expenditures in several countries of the region at the end of the decade.

---

[245] In Poland, for example, social insurance contributions are waived for 'very poor' farmers, people in receipt of unemployment and welfare benefits, serving soldiers and army veterans. The state also provides free medical treatment to children of up to seven years of age, pregnant women, the mentally ill, accident victims and regular blood-donors. In addition, free treatment is provided for alcoholism, drug dependence, psychiatric illness and communicable diseases, such as tuberculosis and AIDS.

[246] To compute the overall tax burden and tax wedge on wages, however, the health contribution earmarked for health has to be summed to the contribution for pensions, sickness and unemployment benefits. Overall, the payroll tax rate is in the range 25-40%.

Table 7.7:Per capita GDP and Health Expenditure in 2000 for a Selected Group of FSE (US $)

	Per capita GDP	Total	Public	Private	Private/Total (%)
Albania	1,112	30	22	8	25.9
Armenia	503	45	18	27	60.3
Azerbaijan	660	10	7	3	32.5
Belarus	2,985	168	137	30	18.1
Bulgaria	1,461	60	57	3	5.7
Croatia	4,263	484	405	79	16.4
Czech Republic	4,940	356	326	30	8.5
Estonia	3,445	211	176	35	16.6
Georgia	**556**	**16**	**4**	**12**	**73**
Hungary	4,532	308	236	72	23.5
Kazakhstan	1,221	69	33	36	51.9
Kyrgyz Republic	268	12	6	6	50.5
Latvia	2,941	191	118	73	38.3
Lithuania	3,059	190	144	46	24.2
Macedonia, FYR	1,768	111	94	17	15.4
Moldova	300	18	9	9	50.7
Poland	4,081	255	192	64	24.9
Romania	1,635	92	62	30	32.6
Russian federation	1,718	109	79	30	27.8
Slovak Republic	3,544	256	202	54	21.2
Slovenia	9,131	706	621	85	12
Ukraine	636	28	18	9	33.3
Uzbekistan	314	15	11	5	30.6

Source: Own computations based on World Development Indicators, 2002.

As Table 7.8 indicates, in a few FSU countries such as Georgia and Moldova, the amount of government health expenditure has been reduced to negligible levels, insufficient to maintain even basic public health services. Also note that all the countries[247] that increased level of real public spending over the decade belong to the Central and East European region or to the Baltic. The "best" performer is Estonia, which increased public health expenditure by over fifty percent in real terms in a period when income fell by 25%.

---

[247] Belarus, not shown in the Table, is the only exception. Belarus has increased real public health spending by 132% over the period 1990-1997. Belarus has not yet undertaken the transition to a market economy.

7.4 The growth of OOP payments and of private providers

As a result of the fiscal collapse of the decade of the 1990s, FSU government health systems experienced drastic reductions in wage and capital investment expenditures. Increasingly, out-of-pocket payments substituted for public funds particularly for drug expenditure and for wages. Physicians tried to make up for the reductions of their official salary payments by offering their services for private payment, typically doing so from the public facilities in which they had previously been employed (see next Chapter).

As table 7.8 indicates, in almost all FSE, formal and informal[248] direct out-of-pocket payments by patients have become one of the more significant sources of funding for health services, particularly in Commonwealth of Independent States countries. In fact, informal payments to doctors and nurses were ingrained in the tradition of the FSE, dating back to the communist times[249], but they do not seem to have been a matter a concern until the last decade.

At the same time, there has also been a growth of the private sector in service provision, so far mainly concentrated in primary and in outpatient specialist care. In Krakow, Poland, for example, a study by Berman and Chawla (1999) shows that over one third of physician contacts (and over half in terms of value) for ambulatory care services take place in private facilities, while 1,096 publicly employed specialists spend an average of 10.8 hours a week working in their private practice. In Hungary, by 1996 80% of the budget for kidney dialysis was paid to private providers (Kornai and Eggleston, 2000). The importance of the private sector is likely to be even greater in those countries of the FSU where the financial collapse of the public sector has been more pronounced, although we lack precise data on the extent of the private delivery sector in this region. Also, in several countries the distinction between public and private institutions is increasingly blurred, as public facilities are managed independently, receive most of their funding from OPP (formal and informal), and their juridical status is undetermined. In fact, some countries (see Georgia, Chapter 8) are struggling to articulate precisely the legal status of facilities that were formally public and that are increasingly *de facto* managed as private facilities.

---

[248] Out-of-pocket payments refer to those payments for health services, which are borne directly by the patients, or their households (for a throughout discussion of the meaning of formal and informal OPP, refer to Chapter 8).

[249] The existence of informal payments was first documented in International publications by World Bank studies early in the '90s. Such studies estimated that 25% and 20% of the funds for health services respectively in Romania and Hungary were out-of-pocket payments and gratuities (WDR, 1993).

7.5 Impact of the growth of OOP payments

A growing body of evidence has looked at these high levels of OOPs and informal payments (see World Bank, 2000a and Lewis, 2001; for Hungary, see Gal, 1999 and for Poland, see Shahriari, Belli, et al., 2002), in an effort to measure their precise extent, and their impact on accessibility and quality of services. Abel-Smith and Falkingham (1996) found that in the Kyrgyz Republic 32% for of respondents cash income was not sufficient to pay for health care. In one-fifth of cases the informal cost of inpatient care exceeded the total cash income of all household members for that month. In Georgia, Gotsadze et al. (2001, p. 37) showed that on average a single episode of hospitalisation in Tbilisi costs GEL 879, which, at the exchange rate of 2.1 GEL per dollar, is equivalent to US $ 418. This sum is far beyond the average monthly household salary in the country, equal to GEL 82.8 (see Chapter 8). In Kazakhstan, according to the 1996 Living Standard Survey (Sari, Langenbrunner and Lewis, 1999), inpatient services cost on average 2 and ½ times the poorest income group's monthly income. Reports of inability to afford care in specific circumstances, or need to sell subsistence produce or assets to pay for health care, show that the equity impact of health payments may indeed be quite relevant in the FSU countries. In 1997, in the Russian Federation 41% of those interviewed reported they could not afford drugs (50% among the lowest income group) and 13% could not afford medical treatment. Among the lowest income group, 36% had to forgo outpatient visits. In Georgia, according to the Socio-Demographic Survey 2002, only 46% of those who were sick sought professional treatment. 20% of the individuals that self-treated did so because they could not afford professional treatment. In 1999 in Tajikistan, among those who needed health assistance, 32% reported they did not seek such assistance because it was unaffordable (Falkingham and Hemming, 1999). The proportion was 42% for the poorest quintile of the population. In this group, 63% were unable to buy prescribed pharmaceuticals, and 21.5% did not have any consultation or prenatal care during pregnancy, because they were too expensive. One third of respondents borrowed money in order to afford health care. In the Kyrgiz Republic, only 41% of those reporting ill health in the bottom quintile sought care, while 62% of the top quintile did. One in three inpatients borrowed money and in the rural areas 45% sold produce or livestock. 15% of the poorest did not seek care even for acute illness because of cost of treatment (Abel-Smith and Falkingham, 1996). In Moldova, more than 30% of respondents in the lowest income quintile reported having to borrow funds to meet the unexpected costs of illness (UNICEF, 1997).

There is not enough evidence yet to assess for which type of services OOP payments are more commonly made (Lewis, 2001). The only consistent evidence is that informal "envelope" payments to physicians constitute the major component of payments for inpatient stays, whereas formal OPP prevail for pharmaceutical expenditure, only partially subsidized by governments, and for outpatient specialist services, where private providers have become more widespread.

7.6 Causes of Informal Payments

The most important single factor in explaining the extent of informal payments seems to be the level of doctors' and health employees' salary (see Chapter 8). The few FSE countries where the existing evidence suggests that informal payments are not widespread, such as the Czech Republic, are also those where physician salaries have kept pace with average earnings during the transition. Where wages are low or even unpaid and the private sector is still underdeveloped, informal payments become the only revenue source for doctors. In turn, low salaries are a consequence of the poor fiscal performance and the excessive capacity in the delivery system.

Apart from the level of salaries, the phenomenon has other complex and deep-rooted causes. In countries of Central and Eastern Europe, and more in general where government financing of health services has not completely collapsed, informal OOP payments tend to be concentrated in a few specific wards and for specific services (for instance, surgeries; see Belli et al, 2002), which are more in short-supply (for example, surgical services). In these countries informal payments can be seen as signalling devices, which indicate where the greater discrepancies are between the priorities assigned through planning and patients' preferences. Thus, they signal where resources are relatively scarcer, and conversely where they are in over-supply. If factors of production could shift more flexibly, the informal payments would be temporary and they could move around within the system depending on where relative scarcities are at any one time. If, as it is, factors are rigid, though, the signalling value of the informal payments is frustrated and the payments become pure rents that can persist indefinitely. By contrast, in countries of the Former Soviet Union, OOPs seem to be much more widespread, and they have de facto become the major source of revenue for the health sector. In fact, a high level of informal payments can contribute to maintain the existing unbalances, as no strong interest group is likely to demand a real change. First, there is no incentive for governments to prevent informal OOP payments, since it allows them to save on labour costs: doctors can supplement their low income by charging patients a fee. Second, most people accept the idea that

the only way to get quality services is to pay directly, so they do not protest. Finally health employees and particularly physicians benefit from OOP informal payments because they earn extra non-taxed income and they can stay in public facilities where they free ride on expensive equipment. Those who are most negatively affected by OOP payments are the poorer and more vulnerable groups, as indicated in the previous section, but these groups are too weak and disorganized to demand a real change.

7.7 Health Sector Reforms in FSE

During the '90s, while undergoing the process of transition to market economies, several FSE began to discuss and to implement health reforms in the finance and the organisation of health services. In fact, FSE can be divided into three groups according to the pace and the comprehensiveness of the health reforms introduced over the last decade. At one extreme of the spectrum we find the "committed reformers" countries (Czech Republic, Estonia, Poland, Georgia), where comprehensive health reforms were implemented. At the other extreme one finds the "resilient non reformers" countries (Ukraine, Slovenia, Bulgaria, Azerbaijan, Belarus) that fixed problems with minimal adjustments and only when they became critical. In between, there is a third group including the majority of FSE, where governments changed parts of their health systems, without touching other key components. Within this group one can also distinguish between smaller countries, such as Latvia, Hungary and Armenia, which adopted a nation-wide approach to reforms, and larger countries, such as the Russian Federation, where reforms were led at the regional level.

The general direction of these reforms has been towards new funding, resource allocation, and provider payment mechanisms, greater autonomy for hospitals, and the establishment of semi-independent sickness funds. The health reforms in FSE shared core features with the reforms implemented in several countries other in West Europe, and other lower and middle income countries, starting in 1990 in the UK. We already discussed in depth the justification, the main characteristics, and the impact of such reforms in previous chapters, and will not repeat the same analysis here (see Chapter 1, Chapter 4, 5 and 6). Here, we would like to emphasize a few elements that were specific to the FSE countries. First, the social and economic transformation unleashed by the socialist system's collapse was highly disruptive and it placed an enormous pressure on health

services[250], making it impossible to maintain the previous standards of accessibility and quality. The dramatic situation that prevailed in some of the poorer countries in the first transition years was thus described by Wiley and Langenbrunner: "During the transition, drastically worsening hygienic norms, shortages of water and electricity, lack of heating caused a daily battle for survival for many hospitals, with rural areas being in a worse condition. The supplies of pharmaceuticals and medical consumables were considerably reduced. In many cases, humanitarian assistance could only partially cover the supply for emergency care, vaccines, and antibiotics" (Wiley and Langenbrunner, 2002).[251] Moreover, the crisis entailed some severe long-term consequences. For example, capital investment was drastically reduced and, as a result, premises lacked maintenance and refurbishment. Thus, the transition years showed once again that the health sector is part of society and of the whole economic system, and it is illusionary to believe that in a context characterized by rapid economic recession and social degradation such as that experienced by several FSU countries, it can stand out as an island of "success" in a sea of despair.

Second, the experience in the region during the last decade seems to suggest that successful reforms inspired by the "managed competition" or "internal market" organization model would comprise a package of several key elements able to sustain and reinforce each other, among which are the following: 1) well functioning primary care system, able to perform its core "gate-keeping" function; 2) a sound and sophisticated hospital management system to cope with the new payment systems; 3) fairness, and independent management on the part of sickness funds; 4) fair and transparent competition among different public and private and public providers; 5) an impartial stewardship role on the part of government.. All of these factors, which could be summarized as: "strong institutional capacity and good governance" were absent or near absent in most FSE countries. As a result, in many instances health reforms in FSE introduced some positive innovations in the health sector, but these innovations were not implemented properly, or were not accompanied by other key components that could sustain them. For instance, the new case-based payment systems for hospitals and capitation-based payment for GPs in general represented a positive change from the previous input-based criteria, but they were not accompanied by new management and information systems able to prevent abuse, and by control over other sources of funding such as those coming from informal payments (see Chapter 8, for Georgia; also Case Study on Hungary, Chapter 6).

---

[250] Again, countries in Central Europe such as Poland and the Czech Republic, are somewhat an exception, as they were able, shortly after the collapse of the communist regimes, to set the basis for the new democratic institutions and for economic development

7.8 The way forward

A first consideration is that any particular reform plan will have to consider and adapt to individual countries' circumstances. Yet, the FSEs also inherited health systems that share core features and problems, and some of the key necessary steps ahead in the reform process are thus the same. All of them are characterized by high rates of prevalence of chronic diseases, mainly cardiovascular diseases, particularly among the adult male population. All of them must find innovative ways to fund their health system within the framework of a market economy. All of them are characterized by obsolete and resource-intensive diagnostic and therapeutic methods, as well as by an excessively heavy and rigid delivery system. In spite of ten or fifteen years of health reforms, all of them to various degrees are still characterized by poor incentives for health workers and facilities within the public sector, and by a yet unclear repartition of roles between the public and the private sectors. The suggested key components of a strategy to move forward are the following:

1. Strengthen Public Health Measures, particularly in the areas of Non-Communicable Diseases and Reproductive Health

The evidence on the burden of disease shows that a few well-focused health policy initiatives in the area of non-communicable diseases' prevention could lead to significant reductions of the burden of disease in the region. FSE governments need to do against chronic disease what they were able to do against communicable diseases in the '50s and '60s, when, owing to strong public health measures such as compulsory childhood immunization, they achieved better health outcomes than other countries with similar levels of income. Health promotion and disease prevention information and communication strategies need to influence life-styles and consumption patterns. These are partly ingrained in deeply rooted cultural attitudes that are difficult to change, but partly they are determined by the level of "health awareness" in the population. For the majority of the population in FSE there is ample scope for improvement, even in the short-term and through simple interventions. The most effective attacking point for influencing life-styles is of course education of the new generation. Healthier life-styles may be promoted through an array of initiatives, ranging from basic hygiene and nutritional notions for kindergarten and elementary school children, to reproductive health courses for adolescents. For the adult population, probably the most effective interventions would be health information and promotion campaigns to be channelled through the media.

---

[251] According to Palu (1999) in the Baltic countries also nosocomial infections became very high.

In the short-term also supply-side interventions are required, such as a stricter control over the most lethal threats to FSE population's health, tobacco and alcohol[252], or health motivated restrictions meant to improve the quality of the air and the water, and to strengthen safety on the roads, in the work place, and in people's homes.[253] As important, health policies should aim at providing families with access to a minimum set of assets[254] and of services, independently from their socio-economic-status. Basic preventive and treatment health services should be part of this set of services.

Finally, reproductive and maternal health is a critical area yet to be addressed, especially in the poorer countries of the region. The potential impact of relatively inexpensive health interventions in this area is immense. Most of the deaths could easily be avoided by simple measures, such as better information and better hygiene in the operating theatres. A striking example is abortion. Education and information campaigns on the demand side, greater availability of birth-control methods on the supply side, could all play a tremendous role in reducing the incidence of abortions, with negligible additional cost for the health system as a whole. In fact, alongside the huge human suffering saved, also in monetary terms health systems would most probably benefit from the diffusion of modern contraceptive methods.

2. Strengthen Health Financing

The evidence presented above clearly shows that the health financing issues facing the two groups of countries, on the one hand the CEE and, on the other, most FSE countries, are of different nature and order of magnitude. In the FSU, with the exception of the Baltic countries, the payroll tax earmarked for health revenues have been negligible (see Langenbrunner and Wiley, 1999). The "extra-budgetary" status of payroll health contributions has made their collection problematic, and tax evasion has been rampant. Moreover, in FSU countries only a fraction of the general budget assignments to health approved by the Parliament at the beginning of each year have actually been

---

[252] Short-term measures should include raising the excise taxes on tobacco or alcohol products, as well as further restrictions such as banning tobacco and vodka advertising. In only a small group of countries in Central Europe, the Czech Republic, Slovakia, Slovenia, preventive campaigns against tobacco and alcohol consumption and for better lifestyle have been fully put into effect, while in a few others, such as Poland, Hungary and Croatia, the first significant steps in the protection of non-smokers have been only initiated.
[253] Among these health-motivated restrictions: 1) stricter regulation of motor traffic speed; 2) compulsory use of seat belts and helmets for travel and work safety; 3) mandatory food supplements in schools; 4) tighter control over food quality.
[254] For example, access to drinkable water, heating and basic energy supplies should be universally guaranteed. During the transition in FSU countries access to these basic services has radically worsened. This has created a

transferred to the health sector by the government over the course of the year[255]. As a result, the resources allocated to health by the government have continued to be largely insufficient, erratic and shrinking over time in real terms. Thus, one of the major health policy challenges in this group of countries is to find alternative health funding sources that are equitable, sustainable, and that do not create excessive distortions in the economy.

In the majority of CEE region, on the contrary, health funding has been more stable and even increasing in real terms in a few countries, such as in the Czech Republic and in Poland. Presently, public health funding predominantly consists of payroll tax contributions. The local and central governments' subsidies finance public health services, current expenditure of the non-contributing segments of the population, and capital expenditure. The rest of the health financial resources originate from patients' direct out-of-pocket payments (OOP). In the Czech Republic, for example, 10-15 percent of total health expenditure is financed by general and local budget transfers. Another 9 percent of the total consists of OOP, and the rest (approximately 80 percent) is funded through the payroll tax. In Poland Hungary and Romania, the share of OOP is larger (approximately 23, 25 and 33 percent respectively).

So, while for several FSU countries the main problem is how to increase the share of government health expenditure, for most CEE and the Baltic countries the main problem is to contain total government health spending, and at the same time to preserve equity and quality of services.

However, one common key challenge faces both groups of countries: how to strengthen health financing sources alternative to the payroll tax contributions. The latter does not seem viable as an exclusive nor a principal source of funding for the health sector in the medium-long term: in some CEE countries the rates of the payroll tax earmarked for health have reached extremely high levels, and yet its revenue is still insufficient, whilst in most FSU countries the revenue from the payroll tax is negligible and insufficient to fund even a minimum level of health entitlements. At a deeper level, the payroll tax would be practical in economies with high shares of formal employment and with relative economic stability, while the evolution of FSE points in exactly the opposite direction. The share of the economy and of employment that are "informal" is continuously rising, also because high payroll taxes in CEE countries have contributed to encourage enterprises to stay "underground".

---

severe health impact, especially for the most vulnerable groups (elderly and children).
[255] Whenever revenues could not match the amount forecasted, the Ministry of Finance set a limit on the cash transfers to each budget line lower than the amount assigned by the budget approved by Parliament.

Moreover, in international perspective, CEE and FSU economies largely base their comparative advantage on the availability of relatively inexpensive but at the same time well-trained labour. Thus, the potential growth-thwarting impact of high health and social security contributions exclusively linked to wages is likely to become a more important factor to consider as the process of globalisation and of economic integration within the European Union advances.

Alternative means of financing health services on a pre-payment basis that could and should be strengthened consist in general budget transfers for the public component, and in voluntary insurance contributions for the private component. However, there is no silver bullet, as these two sources of funding for health care alternative to the payroll tax are also extremely constrained. The government's spending capacity through the general budget is ultimately determined by its fiscal and tax collection potential, which are still very poor in most FSU countries, particularly the direct tax component. It is important to consider that in socialist economies the personal income tax was not a significant source of revenue, because government expenditure was mainly funded through revenues from government owned enterprises (for example, oil exporting public companies), or through indirect taxes. Social services, including health, were directly funded by productive units (Kolkhoz, firms, etc.). It will take time before a well functioning fiscal system based on direct taxation is in place. In turn, voluntary insurance schemes are currently limited to a small wealthier minority segment of the population, and any plan to strengthen them would encounter several difficulties. First, the existing qualitative evidence from the FSU shows that, because of financial scandals and collapses of the past decade, people in general do not trust private for-profit financial or risk-management institutions (see Belli et al., 2002). Second, adverse selection and moral hazard would potentially thwart the growth of any voluntary scheme, unless the government guarantees universality and continuity of enrolment.[256] Despite all the above difficulties, voluntary health insurance schemes still seem to have a great potential to grow in several of the former socialist economies because there is a large unmatched demand for health insurance, particularly against catastrophic events (see Chapter 8).

Finally, to contain the phenomenon of informal payments and reduce their negative impact a multiplicity of interventions is required (see following Chapter), including the strengthening of pre-payment and insurance schemes. Among the other interventions mentioned in the literature are the following:

---

[256] On this point, see Chapter 2.

- Align the scope and scale of health entitlements and of the public delivery system with the available amount of resources. Guarantee a sustainable level of funding for public services and an adequate level of payment for health employees.

- Establish socially acceptable policies concerning formal user charges.

- Improve patients' information and awareness concerning their health entitlements; empower patients' representative groups.

- Reform the criteria according to which health personnel are paid, providing additional financial resources linked to performance, and move away from fee-for-service PS.

- Strengthen transparency and accountability at the different levels of government and within the delivery system. "Objective" resource allocation criteria are preferable to discretionary ones, particularly in contexts characterized by widespread corruption (see Chapter 6). [257]

3. Further Downsize the Delivery System

The "soft" restructuring policies pursued over the last decade, based on the slow financial strangling of under-used facilities through the use of output-based payment systems, have not achieved the expected results.[258] The capacity of the delivery system has indeed been reduced, but extremely slowly and not along all dimensions, and excess capacity still prevents the possibility to adequately finance all existing facilities and personnel.[259] Several countries (including Latvia, Kazakhstan, Kyrgyz Republic, some Oblasts in the Russian Federation and Georgia) are currently moving towards more direct supply re-planning and restructuring policies, managed by the central government (see following Case study). The new official policy documents, known as "hospital restructuring or master plans", outline courageous projects for downsizing the delivery system by: (1) closing idle or only partially operational facilities; (2) relocating and merging numerous single specialty facilities under general hospitals; (3) substituting acute and sub-acute care beds with day-care, rehabilitation and long-term care beds; (4) privatisation of non-strategic facilities and utilization of the revenues for renewal of a few strategic ones. To our knowledge no rigorous evaluation of the results so far achieved in restructuring the delivery system has been published. Anecdotal

[257] For instance, if services for vulnerable groups are for free, it is necessary to specify precisely who the "vulnerable groups" are. Otherwise, as it is presently the case in Georgia, most resources (approximately 2/3) of the program supposedly for vulnerable groups are in fact used to pay for expensive treatments in favour of the Ministers' *entourage* and their families (see following chapter).

[258] Adjusting capacity at the margin (by closing beds) has not worked either.

[259] In many cases, the amount of financial resources received by under-utilised facilities has been just sufficient to keep them alive, but insufficient to unable them to renew their equipment and to provide reliable and timely services.

information suggests that the plans are encountering political difficulties and opposition, and are not yet fully implemented.

Case study 7.2: Hospital restructuring plans

In Kazakhstan reductions in beds and facilities have been in the order of 40-50 percent. Between 1994 and 1997 in the Dzheskasgan region, the organizational and management reforms led to a decrease in the number of beds from 6,225 to 2,919 and to the reduction in the number of hospitals from 55 to 22 (Horst, 1998).

Georgia and some regions within the Russian Federation (Tver, Kaluga, Novgorod Oblast and Chuvash Autonomous Republic) have recently implemented hospital-restructuring plans that originally foresaw a profound downsizing of the delivery system through consolidation and privation of facilities. However, few of the planned hospital closures have actually been completed. One of the more contentious aspects of the restructuring plans concerns the problem of labour force "re-deployment". The health sector is a major employer in the region and physicians representative groups are politically very influential. In order not to be paralysed by physicians' opposition and to succeed, restructuring plans ought to utilize severance pay schemes and other financial incentives for retirement or re-qualification.

4 New Payment System Reforms

The first generation of payment systems' reform was characterized by the introduction of activity-based payments system, fee-for-service or, more frequently, fixed price. As we explained, such payments systems led to some reduction in unit costs, but also to a proliferation of the number of cases treated, and to an overall increase in hospital costs.

More recently, there has been a shift from activity-oriented payment systems to global or capped budgets and capitation funding (Croatia, Slovak and Czech Republic, Kaluga, Tver and Kemerovo region in the Russian Federation; see following Case study), as a means to achieve a tighter control over hospital costs. It is now interesting to monitor this second generation of payment system reforms, such as the Koprevnica pilots in Croatia and the global budgets payments for hospitals in the Czech Republic. Adopting a prospectively-determined global budget, or a capitation funding to achieve more stringent cost-control may induce providers to under-treat or dump costly patients, particularly if payment is unrelated to any performance measure.[260]

---

[260] A separate issue concerns the necessity to align individual incentives with institutional incentives (for Georgia, see Chapter 8).

Case study 7.3: The Kaluga and Tver Regions in the Russian Federation. A recent successful experience of payment system reform

In the Kaluga and Tver Regions regional governments recently introduced global budgets for a pilot group of hospitals and a partial GP Fund holding scheme as an attempt to reduce volume of admissions that had been soaring under the previous cost-per-case reimbursement system. 70-80 percent of the total health budget was assigned to hospitals on a global budget basis, and hospital directors were granted greater autonomy on the use of these funds. The changes have been accompanied by other minor organizational reforms strengthening the power of administration and their control over costs. Early results suggest that:

➢ cost per patient significantly decreased (15-30 percent);

➢ number of inpatient and outpatient specialist visits dropped;

➢ number of beds was reduced by 40% and staff was also significantly reduced.

7.9 Conclusions

Restructuring the health systems of FSE so that they can adequately serve all the residents still represents a major challenge. All but a few of those countries are still struggling to re-shape core parts of their health systems, in financing, definition of priorities for public funding, re-organization of services, human resource training. As Western industrial countries with less radical health reform agendas and within a much more facilitating social and economic environment have found, such reforms are economically institutionally and politically difficult. However, it must also be clear that ultimately for FSE countries the choice is not between reforming their health sectors and leave them unchanged. Fiscal and other pressures make sect oral change inevitable, removing the option of leaving things as they are. The real choice is whether policymakers will allow change to occur haphazardly or will instead try to orient it rationally.

Chapter 8: Out-of-Pocket and Informal Payments in the health sector: evidence from Georgia[261]

8.1 Introduction

"The results of the health reforms are terrible. It left patient without doctors and doctors are left without money to live".

<div align="right">

Focus Group Participant

</div>

Georgia has a population of 4.4 million in a geo-graphical area of 70,000 km^2, bounded by the Black Sea, Russia, Azerbaijan, Armenia and Turkey. Since the independence of the country from the Soviet Union on 9th April 1991, Georgia appeared to be a relatively well-off Republic with promising potential for growth. Instead, soon after independence the country suffered from a period of political turmoil and civil war, which resulted in the influx of approximately 350,000 displaced people from the Abkhazia region near the Black Sea coastline. It also experienced an extremely severe economic decline, caused by the disruption of trade within the former Soviet Union (FSU), and by the difficulties encountered in the transition to a market economy. As a result, over the period 1991–1995 economic output fell by up to 75% (see Chapter7, Table 7.5). Moreover, by the end of the 1993, annual inflation had reached 8400%.[262]

The economic decline and the political and institutional turmoil in the first year immediately following the collapse of the Soviet Union produced a fiscal collapse, and this in turn produced an immensely negative impact on the health care system of the country. Public expenditure on health was reduced to US$ 0.81 per person per annum in 1994.[263] The raising levels of poverty also directly affected the health status of the population, particularly of those that fell into unemployment. IMR, MMR and other priority health outcome indicators worsened dramatically (see Chapter 7).

---

[261] This study was commissioned to me by the World Bank in June 2000, as part of a multi-country research project involving Poland and Kazakhstan. To conduct the study, I devised the questionnaires, I hired local consultants (PMI and Curatio Consulting Company), trained them in conducting in-depth interviews and focus groups in June 2001, and then followed the data collection and analysis. I am fully responsible for the final report, submitted in Summer 2002 to the World Bank.
[262] Both J. Health care financing strategy in Georgia 2002–2007. Tbilisi: TNO; 2002.

[263] World Bank. Georgia poverty and income distribution. Report no. 19348–GE, vol. 1. Washington, DC:

As explained in the previous Chapter, the health care delivery system in Georgia was characterized by a structure known in the literature as "Semashko" model, common to all former Soviet Union countries and based on a capillary territorial network of hospitals and polyclinics. It was a centralized delivery system, which grew over time, during the 1970s outnumbering in terms of beds, facilities, physicians, and nurses per capita, the health systems of the wealthier Western European countries (see Table 7.3 and 7.4 Chapter 7). Financing from the government was primarily based on inputs (bed-days, human resources, equipment, etc.) and resources were used extremely inefficiently.

The economic and fiscal collapse of the first transition years made it impossible to continue to financially sustain the government delivery system.

In 1995-96, in parallel with a new program of macroeconomic and fiscal adjustment, the Government of Georgia decided to introduce major health sector reforms. These included (Gamkrelidze et al, 2002):

1. Removal of entitlement to free health care from the Constitution that was enacted in 1996.[264] Public insurance coverage was limited to services included in a basic benefit package (BBP).

2. Introduction of a social insurance contribution, earmarked to fund health services, set at a rate equal to 4% (3% being employers' share and 1% employees'). Any shortfall in the revenue from this social insurance contribution would be compensated by direct transfers from the general budget (see Box 8.1).

Box. 8.1: Different sources of revenue, and total spent by the government on health

a. General budget: In 1999, health expenditure was equal to 3.6% of total central budget expenditure (total expenditure on health care was equal to Gel 33.2 million, and total central budget spending was equal to Gel 904.8 million) [265].

b. Payroll taxes: Georgia introduced social health insurance (SHI) in 1995, and a semi-autonomous organization -State Medical Insurance Company (SMIC) was created to administer the funds collected for health through the payroll tax, and to set contract for services with health care providers. Participation in SMIC

World Bank; 1999. p. 9

[264] GoG officially introduced out-of-pocket payments in 1996 and even amended the first constitution drafted. The constitutional right to free health services was removed.

[265] Center for medical statistics and information (CMSI). Healthcare in Georgia, 1999: statistical yearbook. Tbilisi; 2000.

is mandatory for all formal sector employees from the public and private sectors. Revenues for SMIC come partially from employees (who pay 1% of their salaries), and from employers (who contribute 3% of the formal wage bill). The shortfalls for payroll tax are matched with transfers from general budget.

c. Allocations from municipalities: Municipalities are the third main source of public financing for health care services. Note that the Ministry of Labour, Health and Social Affairs (MoLHSA) would mandate that municipalities spent on health care at least Gel 2.5 per capita per year (or 10% of their budget, whichever is greater), but most of them do not comply with this regulation. On average they spend approximately 4.9% of their total expenditure on health care (in 1999, total spending by local governments was equal to 320.1 million Gel, and their expenditure on health care was 16 million, according to Kenneth, 2002), with great variability across municipalities. Some of them, such as Tbilisi, spend almost 10%, while others, such as Kazbegi, spend less than 1% of their total on health care (Preker et al, 2002).

Overall, the three above public sources of revenue for health care services were equal to 1.38 of GDP in 2001, as figure 8.1 illustrates.

Figure 8.1: Public and Private Shares of Total Health Expenditure as % of GDP
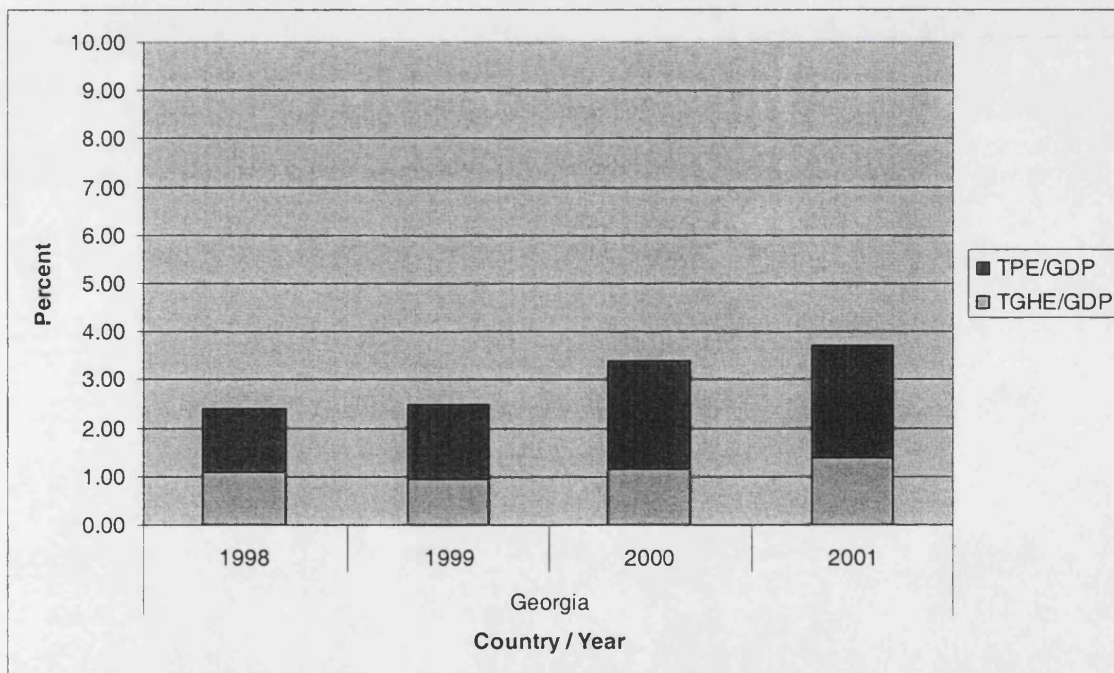
TPE= Total private health expenditure

TGHE=Total government health expenditure

Source: WHO Health for All Database, 2003

3. Redefinition of central and local governments' responsibilities for health by distinguishing between the central and the municipal portions of the BBP. The role of regions and municipalities

was significantly expanded both in health financing and in the management of service provision.

4. Separation of purchasers and providers. Purchasing responsibilities were transferred to the State Medical Insurance Company, to the Municipal Health Funds[266] and to the Public Health Department (PHD). [267] At the same time, hospitals and other service providers were transformed into autonomous juridical entities. They were asked to negotiate service delivery agreements with the public purchasers and would be paid accordingly, on a cost-per case basis for services that were part of the BBP, and from patients' direct payments for all the other services. All health care personnel were removed from the civil service and from the government's payroll. For all the services not part of the BBP providers were allowed to charge patients on the basis of approved fee-for-service (FFS) price lists, and were given the right to retain all generated revenue.

8.2. The extent of out-of-pocket payments in Georgia

With the 1995 reforms OOPs were officially 'introduced' as a source of funding for health services, both as a reward to private practice, as well as in the form of co-payments and user-fees. In fact, informal OOPs were most probably existing during the Soviet times, but with the economic and fiscal collapse of the early 1990s they became much more prevalent. Currently, according to recent estimates, OOPs (formal and informal) are the major source of health financing, equal to approximately 2.3 percent of GDP and 75% of total health expenditure (THE) (see Figure 8.1 above).

As a result of these high OOP payments, health care services and particularly inpatient hospital care services have become extremely expensive. For example, Gotsadze et al. (2001, p. 37) found that on average a single episode of hospitalisation in Tbilisi costs GEL 879 (equivalent to US $ 418), approximately equal to 10 times the monthly average salary in the country.

Moreover, it is estimated that almost half of the total revenue from OOPs is informally paid (Georgia State Department of Statistics, 2000). Georgia is not an exception in the region. The same trend is observed in several other former socialist countries (see Chapter 7). Private OOP spending is

---

[266] Municipal Funds contribute, alongside SMIC, to pay for services included in the BBP.
[267] The PHD is part of the Ministry of Health, and it is responsible for administering public health programmes (e.g., disease surveillance, immunization, etc.).

estimated to be equal to 82% of THE in Azerbaijan, 80% in Moldova and 60% in Kyrgyz Republic (Balabanova and McKee, 2002), which are comparable to the poorest countries of South Asia and Sub-Saharan Africa (see Chapter 6)

8.3 The Study Objectives

The phenomenon of OPP and particularly of informal payments in former socialist economies has been well documented (see Lewis, 2000). However, there is still very little qualitative research on the process collection, on the determinants of IPs and their impact on providers and on the population.[268] Understanding these issues could inform the policy debate, and contribute to articulate new strategies to address the issue of informal payments. This study is meant to: (a) provide an in-depth analysis of the nature, the process of collection and of distribution of OOPs, and particularly of the informal component; (b) identify and explain some of the key factors that can explain the prevalence of OOPs and IPs as a share of the total health financing; (c) understand their consequences, especially on access.

8.4 The study methodology

This is a qualitative study complemented with the quantitative data presented in the previous Chapter (see State Department of Statistics, 2000, and Mossialos, 2002). The qualitative study considered two target groups: patients and service providers. Patients were engaged in in-depth interviews and focus group discussions, while for providers only in-depth interviews were conducted. Two geographical localities were selected for the study:

1. The capital city of Tbilisi, which houses almost a third of the whole population of Georgia. In addition, the most important healthcare facilities of the country are located in this city: 74 hospitals, 123 polyclinics and 16 dispensaries. The doctor to population ratio is one doctor per 118.6 individuals, and there is one hospital bed per 135 resident, with utilization rates being only 32%. Almost 45% of the total National health spending occurs in Tbilisi (Mossialos et al., 2002).

---

[268] I carried out together with Helen Shahriari one of the few qualitative studies similar to that presented here (Shahriari, Belli and Lewis, 2001). This study shows that the nature, extent and causes of OPP and of informal payments are very different in Georgia and in Poland.

2. A country-representative Rayon (district) –Gori- with its urban and rural locations, where choice of services is limited relative to Tbilisi. Gori is of one of the six Rayons (districts) in the Shida Kartli region. The city of Gori, with a population of up to 60,000, is the administrative and cultural Centre of the district. The whole district houses 181,988 individuals, who are served by four hospitals (including a 90 bed maternity home), with a total bed capacity of 586 beds, 2 rayonal polyclinics and 22 rural ambulatories (village clinics) located in the surrounding villages. There is a doctor for every 398 individuals in the district, and one hospital bed per 310 (Centre for Medical Statistics and Information, 2000).

The principal aim of in-depth interviews with patients was to understand the level and pervasiveness of OOPs, formal and informal, whether or not and to what extent OOPs create a barrier to access health services, and the dynamics of payment. The interviewers asked specific questions on when, how, to whom, and why payments generally occurred. Guides for in-depth interviews included questions concerning illness and non illness related episodes (such as pregnancies or preventive services), services included and not part of the BBP, cases when services were fully utilized, as well as cases when treatment was forgone, or delayed or interrupted because of financial or other reasons (see modules for interviews and focus group discussions in the Annex 8.2).

Focus Group discussions reiterated some of the same questions of the in-depth interviews, but also concentrated on more general policy-related issues concerning service provision and payments. The responses given during the Focus Groups were compared with those given during the in-depth interviews. The objective was to ascertain whether confidentiality (obviously absent in a focus group discussion) made any difference and to further probe the results of the in-depth interviews. Contrary to what was found in Poland (see Shariari and Belli, 2001), in Georgia responses during focus group discussion perfectly matched those during in-depth interviews, suggesting that people are willing to speak openly about their experiences with paying for health services.

In the second part of the Focus Groups, questions about the perceived causes of payment, and households' perception concerning the viability of different policy interventions to contain IPs were discussed.

In-depth interviews with providers aimed at exploring the dynamics in the flow of funds within health facilities, including the mechanisms of payment collection, distribution, and utilization of revenues from OOPs (formal or informal), and health workers' perception of the causes and

consequences of IPs. Pre-testing of the questionnaire indicated that providers were also willing to speak openly about their experiences with IPs.

For the in-depth interviews with households (HHs), a "snow-ball" sampling method (see Denzin, Lincoln, 2000) was used to select those HHs where at least one member had used hospital services or outpatient services over three months prior to the interview. The total number of individuals residing in the HHs that were interviewed was equal to 406, out of which 146 individuals (36%) reported having used health services (because of illness, prevention or because they were pregnant) during the 3 month prior to interview. The sample-selection criteria were also meant to capture the urban/rural differences and the differences between different economic, age and gender groups.

Table 8.1: Sample Distribution for the Focus Groups

	Poor	Rich
Tbilisi	3	3
Gori	3	3
Total	6	6

Table 8.2: Sample Distribution for Providers and Households selected for in-depth interviews

	Providers		Households	
	Hospital staff	Outpatient facilities' staff	Urban	Rural
Tbilisi	21	10	58	
Gori	6	3	15	26
Total	27	13	73	26

The sample of users from urban and rural areas was further disaggregated into sub-sets according to inpatient–outpatient, pregnancy and child delivery, paediatric, and male–female (the detailed breakdown of HHs selected for in-depth interviews is presented in Annex 8.1 on page 267 and subsequent pages). HHs' economic status was estimated through questions concerning total HH monthly expenditure and income. Reported income was related to region/city specific income levels reported by the State Department of Statistics. A specific tool (pre-questionnaire) was developed and administered by the surveyors for the respondent recruitment for the in-depth interview or for focus group discussion.

The pre-questionnaire allowed adequate representation of various groups to be reached. A total of 12 focus groups with 119 individuals, and 186 in-depth interviews were carried out.[269] Obviously, given the limited sample interviewed and the qualitative methods utilized, this study does not aim to provide statistically representative data.

The fieldwork was carried out during June–August 2001. The data analysis was carried out during September–December 2001.

The limited sample size, the impossibility to carry out follow up interviews, and the qualitative methodology utilized limited the scope of the results we found. As previously explained, our aim was to reach a more in-depth understanding of the dynamics, causes and consequences of informal and formal direct payments in the Georgian context. Our study findings complement the evidence from other studies, which included a household health expenditure quantitative household survey carried out in 2000 (see State Department of Statistics, 2000).[270]

8.5 Definition of Out-of-Pocket Payments (OOP) and of Informal Payments.

Out-of-pocket payments (OOPs) refer to those payments for health services that are borne directly by the patients, or their households. The emphasis of this definition is on the direct nature of the payments, which distinguish them from the other major sources of health care financing, private health insurance, compulsory social insurance and taxation. The latter all share the common feature of being pre-payment mechanisms: people pay in advance, according to different criteria (according to their health risk in case of private insurance, or to their ability to pay in case of social insurance or taxes), in order to have their health consumption partially or fully subsidized at the point of service. The informal component of OOPs is defined as the payments (cash or in kind) made to service providers (person or institution) by those people who are entitled to the services, in addition to any legally defined payment.[271] This can take several forms, including direct cash payments to medical

---

[269] 146 in depth interviews were with individual patients, and 40 with providers (30 doctors and 10 nurses). The results of FG were video taped and in-depth interviews recorded on tape recorder.

[270] Some of the results from this quantitative survey are presented in Chapter 7.

[271] In order to facilitate a common understanding of informal payments, we would emphasize the following key elements of this definition Belli and Gaal, 2002):

1. Informality: Informality means that the economic activity is not registered. This is a common feature of all

doctors, gift to nurses, or the in kind provision of certain elements of services, such as drugs, nursing or meals in inpatient care, which should otherwise be the responsibility of the provider.

According to the above definition, given the set of health entitlements in each country, namely the rights to free or subsidized access to health services as defined by law, it is to be considered as formal any payment for those services that are not included, and as informal any payment for services that are included in the set of these health entitlements (or above the legal co-payment, in case of services only partially subsidized by government).

This seemingly straightforward distinction, however, is extremely difficult to apply empirically in the case of Georgia. In theory, the law on health entitlements states that patients have the right to services included in the Basic Benefit Package (BBP) for free or in exchange for low co-payments, and that for all other services every medical provider should determine the prices at the point of service. These prices vary according to type of service and facility, and should be posted in health facilities and publicly known. However, in reality we found that patients are unable to distinguish formal or informal payments, because they do not know what their entitlements are.[272] Thus, the study investigates into all OPP, and not specifically the *informal component*, although some sections

---

types of informal payments. But not all unregistered payments given to service providers are informal payments according to the above definition. We are not interested for instance in those payments that are due to the provider according to a set price for a determined service, but for which the provider does not give a receipt to avoid paying tax.

2. Illegality: A second important defining dimension of informal payments is legality/illegality. Laws and regulations in each country determine the definition of informal payments, by determining entitlements for health services. They define (or should define) who is entitled to which services, in exchange for what (i.e. what could be charged formally for the services). Not all informal payments, though, are illegal. They are not illegal if patients give them spontaneously, as an expression of gratitude or of appreciation for an extra-service received, without any pressure from the provider (such liberal payment can be compared to an extra-tip given to a waiter in a restaurant). On the other hand, it is true that everything that is illegal payment is also informal by definition.

In general, it is empirically impossible to ascertain whether a certain informal payment (a payment for a service to which the patient is entitled for free, or a payment above the legally set fee) is also illegal. Thus, the definition of informal payments in the text, which is the only one empirically viable, also includes legal but unrecorded payments. Moreover, it is easier to ascertain informal payments when patients are entitled for a service completely free of charge. It is more difficult to identify informal payments if a nationally defined formal co-payment can be charged for the service concerned, and it is even more difficult when each provider, as in the Georgian case for all services that are not part of the essential package of services (BBP), can charge a different price for the same services.

[272] First, the population has sometimes a vague idea of their rights to free services, or they do not know the official co-payment rates, nor they know the official prices for the services not included in the BBP. Second, with the exception of the services, which are part of the BBP, it is also impossible to separate formal from informal payments in the phase of data-analysis from households' responses concerning the amount paid for different type of services, because prices vary in different facilities. Furthermore, we discovered that it is sometimes difficult to distinguish between the two also from providers' interviews.

are exclusively dealing with the informal component of payments. From providers' interviews, the study was able to capture some interesting information linking formal and informal OPP[273]. In summary, all the episodes of care where payments occurred investigated in this study can be classified in 5 different categories:

1. Fee-for-service payments for the services that are not part of the BBP. In principle official fee schedules are set by each single medical provider, according to lists reviewed and approved by the MoHLSA licensing department during the process of licensing.

2. Private payments for drugs purchased from pharmacies. Private spending on pharmaceuticals is estimated to be somewhere between 26% and 53% of total private spending on healthcare (Gotsadze et al. 2001, SDS 2000[b], Skarbinksy 1999, Mays and Schaefer 1999). Except for drug-prescriptions within services that are part of the BBP (for instance, pain-killers for terminally-ill patients, or drugs for emergencies)[274], there is no other form of drug provision on a pre-paid basis.

3. Co-payments for services partially covered by private insurers. In Georgia a few private medical insurance companies are operating, although their scope is still extremely limited. During 1998-1999 the share of private health insurance over the total private insurance premium collection rose from 5% to 7.9%. However, in 1999 health premiums totalled only 468,000 GEL (≈ US$ 234,000, ISSS 1999), which in per capita terms is hardly US$ 0.045, and approximately 0.15% of total estimated private spending.

4. Co-payments for those services that are part of BBP, but are not fully subsidised by government. This happens for services, such as emergency services, which belong to the Municipal part of the BBP. For instance, in year 2001, hospital services for adults (15+) that were part of the Municipal BBP were subject to a fifty percent co-payment.

5. Informal payments to physicians and other personnel for services that are part of the BBP and should be provided for free at the point of service.

---

[273] See section: 8.2.1.6: 'How Much Is Paid? Relationship Between the Formal and Informal Components of OPP.
[274] In fact, a minimum share of expenditure within the BBP is allocated to drugs (only 5% of public funds).

8.6 Results

8.6.1 How payments are made

"...There are various forms of payment and various reasons why patients pay informal fees: gifts, money demanded or voluntarily paid. I have not researched into this issue but I am sure all these forms exist in our facility". Hospital Provider Interview, Tbilisi.

Payment occurred in 123 of the 146 cases investigated, and almost universally respondents agreed that paying for health services in Georgia is a really common and accepted practice. The general perception among those interviewed was that, unless a patient had good connections with somebody in the government (local administration) or directly with doctors, s/he had to pay to get treated:: *"You go to a policlinic to get registered and have to pay, you see a doctor and have to pay, call ambulance and have to pay, need the drug and have to pay". FG, Low Income"*

The study also captured several forms of OOPs, as well as various circumstances in which they occur. For services that are not part of the BBP, patients are charged either according to official price lists, or according to informal fees negotiated with providers with approximately equal frequency (see discussion hereafter). For services that are part of the BBP and for which patients should not pay according to the existing laws, payments occur informally, and doctors and nurses use various methods, some "softer" and other more direct and coercive in order to be paid. Patients report that doctors and nurses explicitly complain about the economic hardships they face, or they do not pay adequate attention or are rude to them until they pay, or –more often- explicitly ask them or their family to pay.

When payment is explicitly demanded, it is generally requested ex-ante or during treatment. A significant number of individuals recalled stories where physicians and sometimes nurses demanded payment from the patient as a pre-condition to access treatment or to complete it: *"When you call an ambulance, before they even inquire about the patient's condition, they ask you if you have 10 Gel[275], and they will come only if you have money to pay." FG, Gori.* In few cases, patients felt so directly

---

[275] Gel=Georgian Lari is the Georgian currency. At the time of the study, its value of 1 Gel was approximately 0.5 US$.

forced to pay that they remembered the experience as "extortion": "...They made it clear that if I did not pay I would not be treated. In fact this was simple extortion". FG, Gori.

Money is not only demanded for treatment or drug purchases, but it is sometimes demanded also to provide sickness certificates or exemptions from military service. We encountered cases where providers exploited to their advantage existing laws or regulations that were meant to favour patients. For instance, SMIC issues an "insurance card" for all children who are entitled to free public coverage under the Children 0-3 Program [276]. Providers should distribute the insurance card at no cost whenever they attend to a delivery, or an eligible newborn seeks care at their facility for the first time. However, the study documented cases where parents were charged for the insurance card, and then, despite showing the card to another provider, they also paid for the health services received by their children. A young mother in Gori said: *"I had to pay for the insurance card 25 Gel. The card said that treatment for my child was free. I think the government pays the cost. I was told that if I do not accept the card I had to pay a fine in the amount of 100 US$. Now I have this insurance card, but I still have to pay every time I go to the hospital. Everybody who goes to that facility has the card and everybody has to pay"*. FG, Gori

The study also captured few instances where payment was apparently given in a completely voluntary way, as an expression of gratitude or/and because it is perceived as the "norm": *"It is a habit to pay doctor in order to express thankfulness. As an example my co-worker took the child for vaccination that is free, but she paid money. When I asked why she did so, she responded that she felt uncomfortable not to pay the doctor"*. FG. Tbilisi. Voluntary payments seem to be more prevalent in rural areas, and for outpatient services rather than for inpatient services. *"...We are very pleased with our village doctors. They are doing well; they are knowledgeable and available whenever we need them. We try just to thank them at the best of our ability"*. Household Interview, rural area, Gori district.

Informal charges are paid with cash, or by 'exchanging' goods or services (barter), or by offering in-kind contributions or gifts. All of these forms were found in the study, even if cash payments largely prevailed. Gifts are common when there is a personal relationship between the patient and the physician/nurse. In these cases, offering cash would be embarrassing for the patient, but at the same

---

[276] One of the programs in the BBP

time leaving the doctor's office without expressing gratitude would be culturally not acceptable. Gifts are usually of symbolic value (such as a chocolate box, or cheese, or flowers).

It is also common for households to contribute some of the supplies needed for hospital treatment (in kind contributions), most often medicines and food, but also syringes, bandages, sheets or blankets, and even blood for transfusions. These contributions make up for inputs that are frequently missing within facilities, and households can sometimes find it extremely difficult to find them in the market. Some patients were given blood by other family members and/or their relatives.

8.6.2 How much is paid?

Prices vary according to the facility, the type of service, and providers' assessment of patient's ability to pay. Sometimes price negotiation takes place between providers and patients. The study captured two distinct occasions when such negotiations took place. The most common one is for delivery services. Physicians are approached well ahead of time to secure their availability for the delivery, and the price of their service becomes subject of negotiation. The other case concerns negotiations over the price of inpatient or outpatient curative services. In their in-depth interviews, physicians said that during these negotiations they utilize the official price list that sets a price for each service (which is known to them, but usually not to the patient) as a benchmark for charging. They also seem to be guided by their perception of the patient's ability to pay. These negotiations serve two distinct purposes:

a. To determine prices that will be "affordable" to the patient. A significant number of interviewed providers said that they feel uncomfortable about asking for payment, and try to provide services to the poor at "affordable" prices or even for free. Forgone revenues on treating poorer patients are often compensated by charging wealthier patients more. As a result of this positive price-discrimination, poorer patients seem to be able to receive the same services at significantly lower prices than the better off: "...The price is set after negotiations with the patient and according to her/his ability to pay. However the price does not determine what kind of services the patient will receive". Gori, Provider Interview. In a few cases, also patients from well-off households complained that they were asked to pay more, when doctors suspected they were relatively wealthy: "...I think we had to pay 500 Gel, but they asked us to pay 1,000 Gel. Probably they noticed that we had money and demanded more". Patient Interview, Tbilisi. Note that by reducing price for the poor

and increasing it for the rich doctors enhance access for the poor; however, these cross-subsidies are completely left to each provider's good-will, and each physician's willingness to collect informally is an important factor as well to determine the price each patient is charged. As rightly argued by Ensor and Killingsworth [277]: "The poor are unambiguously subsidized by the better-off in case of explicit cross-subsidies, whilst in the case of informal price discrimination willingness to pay determines contribution. While willingness to pay may often accord with ability to pay, this need not be the case. Those on lower income with fewer opportunities to obtain treatment elsewhere, and an immediate need for medical care, may show a strong willingness to pay to the extent that they borrow or sell goods in order to obtain the necessary funding". Moreover, it is not clear whether the criteria used by providers to judge their patients' wealth are accurate or not. In rural areas doctors know their patients' economic situation, but in urban areas they base their assessment mainly on questions to their patients about their profession and area of residence, and look at their attire.

b. To reach an agreement through which physicians reduce the official part of payment, in exchange for some "informal" income that they can directly pocket. Physicians receive a low share of all the revenue that hospitals collect officially (see below), and at the same time official charges for services outside of the BBP have raised dramatically over the last few years (see for example, State Department of Statistics, 2000). Thus, both physicians and patients have an incentive to 'informalise' their transactions. Physicians can pocket a higher personal income by not reporting their revenues to the facility administration, and patients can receive services at a lower price if the physician does not report her/his service, or manipulates the diagnosis so that their service becomes partially or fully funded under a publicly paid program. Thus, physicians frequently arrange payments in a way that is advantageous for both (patient and doctor): "We were told that officially the surgery would cost 1,200 Gel, and that we would be expected to pay more to various staff. However, the doctor proposed to change the diagnosis, and in that case the government would partially cover the cost of treatment. We would have to pay only 600 Gel to the doctor ". Patient Interview, Tbilisi. A specialist, describing his outpatient activities in a hospital in Tbilisi, said: "Sometimes patients approach doctors directly, when they want to avoid official fees. For example official fee for a doctor consultation is set at 15 Gel and for a professor consultation at 20 Gel. When patients strike a deal with the doctor they pay less (e.g. 5-10 Gel). Of course this payment bypasses the cashier, so payment is not recorded, and the money becomes the doctor's personal income ". Provider, Tbilisi.

---

[277] Ensor T, Killingsworth J, editors. Paying for health: unofficial payments for health care in low and

This finding, namely IPs are not always paid "on the top of" official fees but may at times contribute to reduce the total cost of treatment for patients, is one of the most interesting and new in our research. It is also remarkably different from the findings of similar studies conducted in countries where IPs are not as widespread as in Georgia (in Poland, see Shahriari and Belli, 2001, and in Bulgaria see Balabanova and McKee, 2002) and where authors found that IPs increased the total amount paid by patients.

In the Georgian context, proposing to further raise official fees (see Lewis, 2000) as a means to contain IPs may in fact contribute to further increase the incentive to maintain informal transactions.

8.6.3 Payments for services that are part of the BBP

According to the law, services included in the BBP (see Annex 8.4) should be provided for free (small co-payments exist only for services funded by municipal governments). In reality, our study found that in general patients pay for services that are part of the BBP, although they may report to have received a 'discount' for these services: *"My sister has a child, and she has the SMIC insurance card for 0-3 years old. Her child needed surgery and the cost was 800 GEL. The (public) insurance partially covered the surgery cost, and we had to pay only 300 GEL. Due to the financial assistance received under the SMIC scheme the cost of the surgery was bearable for our family". FG, Tbilisi.*

Furthermore, our findings confirm results from previous studies (Gotsadze et al., 2000) that the population does not precisely know which services are part of the BBP. For instance, a mother who had heard about the program for children 0-3 was still sceptical when she joined the Focus Group: *"My friend told me that in their polyclinic children receive services for free... I do not know if there is any age limitation. I think this is an initiative of their polyclinic". FG, Low-income.*

Thus, our study found that several households had heard in the media messages about free-health services, but that there was a widespread confusion as to what services were part of the BBP, and distrust towards these information campaigns. The messages given were in stark contrast with the daily experience of all respondents, where in fact no service is (almost) ever given for free: *"We hear on the media that services for pregnant mothers and small children are for free. Probably some lab test are free, or maybe something else, but our experience is that when we go to hospital they always charge us". FG, Gori, Low income.*

---

middle-income countries. Cheltenham: Edward Elgar, 2000:8.

For example, all couples that we interviewed who were planning to have children were accumulating savings in order to pay for delivery services, which supposedly they could receive for free in public facilities. A mother, who had given birth in a maternity home in Tbilisi, seemed to know and to accept to have to pay for services in maternity hospitals: *"For the delivery we had to pay the doctor 200 Gel...We agreed the price with the doctor. They also told us that as long as we arrange for their guaranteed assistance during delivery several month prior to delivery, the public coverage does not work, and we are not eligible for free services".* Patient Interview, Tbilisi.

Physicians also admitted to ask for payment to provide BBP services. They justified themselves by complaining that their low official income forced them to do so.

8.6.4 Payment collection and distribution mechanisms

The dynamics of payment collection captured by the study are quite diverse, though some common characteristics can be found. The process of payment also varies by type of ownership. Established private providers only collect payments through cashiers, according to posted prices: *"In private clinics staff immediately informs you about the price of services and only after receiving treatment you pay your dues at the cashier, for which you get receipt".* Provider interview, Tbilisi.

In public facilities [in both localities Tbilisi and Gori] payments are given either to the cashier, or directly to the doctor (nurse). The latter form prevails in outpatient facilities, while the two forms occur in approximately equal proportions in hospitals: *"...Out of 20 such transactions only 10 are recorded on the official books (deposited with cashier), while revenues from the other 10 are shared among various physicians who were involved in the patient's treatment process".* Provider interview, Tbilisi.

Payments not officially recorded are either collected and pocketed by the physicians who treats the patient, or (less frequently) shared among physicians who work in the same ward. Separate and smaller 'side-payments' are then also offered to one or two of the nurses who contributed to the patient's care. In inpatient cases sharing arrangements are more frequent. The study also encountered physicians who claimed that part of the informal cash collected was used to pay for essential non-labour inputs, primarily drugs. These investments seem to be guided by the desire to improve quality of care and thus attract more patients. *"Part of the informal income generated by the medical staff was accumulated in a special fund. The money from the fund was used to do little maintenance*

to the ward and its equipment. Also, it contributed to purchase some stock of most needed medicines." Hospital Provider Interview, Tbilisi. "Our head of department collects money from the staff, which is spent to purchase cleaning materials, lamps, paper and pencils and for the other things that are needed. Otherwise nobody provides these things to our department". Hospital Provider Interview, Gori.

Overall, our study found some sense of unity or solidarity among different medical staff employed in the same ward or the same facility. This is confirmed by another interesting finding: given the scarcity of patients who seek care at facilities, sometimes physicians who work in the same ward agree on informal "rotations", so that part of the staff who is supposed to be on duty on any specific day is in fact absent, substituted by other colleagues. By these means, each physician can expect to be reasonably busy during the time he/she spends in the facility, collect at least some informal revenue, and at the same time have more time to look for other employment opportunities outside the facility: *"We are too many compared to the workload, and our remuneration depends on how many patients we treat. Thus, doctors have an internal agreement, whereby we 'rotate': one week one doctor serves all patients, and the following week another one comes in. The exceptions only exist when a patient asks for a specific doctor". Hospital Provider Interview, Tbilisi.*

8.6.5 Causes of the prevalence of OOP payments

This section specifically focuses on the reasons why OOPs seem to be so widespread in Georgia, and why a large share of payments seems to occur informally. A plurality of factors, which can be grouped into three broad categories, can contribute to the present situation. These three broad categories are: b) demand side factors; b) contextual factors; c) supply side factors.

8.6.5.1 Demand-side factors

Factors linked with patients' tastes, expectations and fears, as well as cultural norms ingrained in the Georgian tradition, contribute to maintain a high level of OOPs and IPs. The two main demand-related factors are patients' willingness to get additional or better-quality services relative to the average level, and to skip long waiting lists. However, unlike other former socialist countries where OOPs are not so widespread, in Georgia these demand side factors do not seem to be critical, and cannot explain why IPs constitute such an important share of the total OOPs. The fact that in Georgia the general attitude among the population is one of surrender to OOPs, seems to be a consequence

rather than a cause of their pervasiveness. Because currently almost everybody pays directly for health, formally or informally, this has become by and large socially acceptable. Other demand-related factors include:

a) <u>Cultural norms and traditions:</u> It seems that directly paying doctors and nurses has been a cultural norm in Georgia for a long time, since socialist times and even before. Several patients admitted that they used to pay providers also during the "old Soviet days", although on a different scale. *"When I do not have money I do not go to a doctor. I will be embarrassed to show up and not be able to pay. We have always paid for health services"*, *Patient interview, Gori.*

b) <u>Gratitude:</u> Several patients were willing to pay something extra as a sign of gratitude: *"...When a doctor saves your life you want to thank them". "...We are paying to express our thankfulness and respect for the doctor's work". Patient Interviews, Tbilisi.*

c) <u>Desire to Support Doctors:</u> In general, the Georgian population believes that health workers are suffering from the same economic hardships as all others after the collapse of the Soviet Union. Thus, people are willing to support them by paying them directly in exchange for medical services: *"Doctors suffer and they do have families that need support. This is the reason they are forced to charge IPs. If they could have higher incomes, I am sure they would never do this". FG, Tbilisi.*

d) <u>Lack of Trust:</u> The vast majority of the respondents as well as of providers interviewed expressed a complete distrust in the government, central and local. *"The government cheated us so many times, that there is no way we can trust them. Just looking at the experience of our savings that disappeared (in banking crisis in the mid '90s) is a sufficient example for this". FG, Gori.* The majority of them did not trust private for-profit insurers either: *"Today it is impossible to trust anybody. Nothing is stable in this country. If you have money you better keep it at home". FG, Low income, Tbilisi. "These companies [private insurance companies] will disappear like the banks did in the past and we will be left with nothing". FG, Well off, Gori.* Thus, people distrust insurance and any other type of pre-payment mechanism; they prefer to pay directly for services, because that is the only way they believe they would receive something in exchange. Such deeply rooted distrust was highlighted by other studies (Bennet et al., 2001).

8.6.5.2 Systemic Factors

The key factors in explaining the prevalence of OOPs and IPs are systemic or contextual factors. These were captured in our interviews with providers and patients, as well as in our discussions with government officials, and are also identified by other studies used as a basis for our investigation. These factors are the following:

a) severe under-funding of health services by government after 1991;

b) poor definition of the benefit package (BBP);

c) overcapacity in the delivery system.

a) All health care providers lamented government's continuous under funding of the health sector during the last decade. In turn, this was due to the economic and fiscal collapse that followed the transition from communism. The government not only ceased any further capital investment, but it could not continue to fully finance essential items in the current expenditure budget, such as medicines and salaries. For instance, a nurse employed in the maternity home in Gori expressed her frustration, shared by several of her colleagues, by saying: *"It has been several years since the government, supposedly due to limited financial resources, has paid in full the amount dues to our facility"*. All providers interviewed in Tbilisi as well as Gori voiced similar complaints. One senior physician thus described the financial situation in his department: *"We have not received our official salary since last August (almost 12 month) due to delayed payments from Tbilisi municipality. Our doctors sometimes do not have money to purchase a loaf of bread"*.

b) The list of services included in the BBP is extremely complex, convoluted, and it has been revised almost every year since it was first introduced in 1996 (see Annex 8,4). As a result, only a minority part of the population, those who are more educated and informed, know which health services should be available for free, and which people should pay for. For example, the study found that, whilst some mothers, mainly from wealthier backgrounds, received immunizations for free, others were paying for these services. A mother from a poor household grieved: *"Immunization that used to be free can now be charged. They demand money for syringe and other things"*. FG Low income. Sometimes even health care professionals did not know precisely which services were part of the BBP, and which were not, or what rates or co-payments they were supposed to charge.

c) Finally, the problem of excessive capacity and resulting under-utilization of the delivery system has contributed to exacerbate the impact of the disastrous fiscal and health financing situation. A thin government budget is spread over a large number of facilities, which are also competing for a limited number of patents with ability to pay. The excessive capacity of the public delivery system has been well documented elsewhere [see Zoidze et al., 1999, and Gotsadze et al. 2000]. The problem is still largely unresolved, in spite of a large downsizing which took place during the last decade (see KPI, 1999). The perception of a severe under-utilization of the delivery system also emerged in our interviews with providers: *"...On average we see 6-8 patients a day. Personally I see 2-3 patients a day. The occupancy rate is not high. We are idle most of the time, even if currently we only have 2 doctors and 4 nurses, half the staff we used to have before." Provider from Hospital in Tbilisi.*

8.6.5.3 Supply-side factors

"Doctors are presently in a very poor condition. Petty traders on the street earn more than doctors." Provider Interview, Gori.

Two supply-related factors emerged as critical in our research to explain the prevalence of IPs. These were:

a) inadequate official income of health personnel, and

b) lack of transparency in the administration and management of provider units.

a) According to formal rules, since 1995 the Government of Georgia does not finance health facilities according to their inputs, nor pays salaries to healthcare workers. Staff is directly employed and paid by each facility, and the latter are paid by government on a cost-per-case basis (see Chapter 4) only for their provision of services that are part of the BBP (such as deliveries or treatment of 0-3 year old children). The main source of revenue for provider facilities are the user fees charged to patients for the services that are not part of BBP. Thus, according to the rules these two sources of revenue (government funds for BBP services and official user fees for all other services) should cover all recurrent costs for health care facilities. In reality, these official sources of funds are insufficient to cover all competing claims, and in addition facility directors or managers seem to enjoy a sort of "monopoly power" over their distribution (see next section). Medical and paramedical

staff is usually employed under output-based contracts, whereby they receive a proportion of the total revenue they are able to generate for their facility (see Chapter 4). Because of lack of funds, corruption and excess supply of physicians and paramedical staff, in fact physicians receive a share equal to only 7-8 percent of the total official revenue they generate, and nurses only 4 percent or below.[278] Given the prevailing activity levels, according to our findings this would translate into an average monthly net income of approximately 55 GEL (≈26.5 US$), well below the average monthly salary in the country, which is GEL 82.8 for formal sector employees. In reality, it is a common practice (and sort of accepted by all within the facility) that doctors would make most of their revenue from IPs. In fact, these are not only paid over and above official fees, but at times "replace" them according to informal agreements reached between patients and physicians (see section 8.5.2). In all our interviews, medical staff complained that their official remuneration was insufficient, and did not adequately reflect their contribution. As a result, most of them admitted that whenever possible they tried not to report the real volume of activities and revenues they are able to generate to their administrators. For example, a surgeon from Tbilisi said: "When I perform surgery and I report it, I have to deposit 300 Gel with the accounting office. These funds should provide income for the anaesthesiologist, for nurses, and for other staff, and pay for all the other costs in the ward. However, I prefer to keep this money and share it with the other staff, rather than to deposit it with the accounting office. The reason is that I only receive 46 Gel for every 1000 Gel I generate for the facility. This is unacceptable". Another doctor in Tbilisi so described the situation: *"Doctors are not pleased that they generate 8 Gel for the facility and get less than 1 Gel as their income. They prefer to receive 3 Gel informally, which is cheaper for patients. The situation forces them to act this way. At the same time everybody knows about it [IPs], but nobody reacts because it serves everybody's interests: the administration's as well as doctors'."* Provider Interview, Tbilisi. The financial situation of nurses is even more difficult than for physicians, because nurses seem less able to make up for their low official income by asking IPs. As Annex 8.3 indicates, the number of nurses has been drastically reduced over the last decade. One of the nurses interviewed said: *"Nurses' income is very low. It could be at around 4% of the unit price of service. My monthly net official income amounts to only 9-15 Gel".* Provider Interview, Tbilisi.

---

[278] This is an extremely low share by any standard in a labour-intensive sector such as health. Note that regulation of contracting with the medical staff varies from one facility to the other, so that the study found high variability in the share of revenues generated that physicians reported to receive in different facilities.

b) The 1995 reforms led to the "corporatisation"[279] of health provider units. In fact, Government of Georgia retained ownership of most provider units, but the latter became incorporated under private law, governed by a Board and managed by an executive director. These took on day to day management of health facilities. In the intentions, self-management would empower provider units to administer resources more effectively, and direct input-based control by government would be substituted by results (outputs and outcomes) based control mechanisms. In reality, the reform failed. Decentralization of power to the provider institutions was not accompanied by efforts to develop better capacity to adequately manage services, and at the same time regulatory/accountability mechanisms based on results or performance remained extremely weak. In addition, even basic principles of sound corporate governance practices have been altogether ignored (Gotsadze et al., 2001) not just in the health sector, but in all sectors of the economy. Thus, in practice the 1995 "corporatisation of provider facilities gave unlimited power to these facilities' executive directors. Not surprisingly, such unlimited power seems to have been frequently abused, according to our study findings. The study found a striking consistency in physicians' and nurses' allegations of non-transparency and of misuse of official funds within their facilities by their directors. First, practically all interviewed medical staff said that funds were administered secretively, and that only directors controlled and knew how official revenues were utilized. For example, a specialist in Gori said *"The administration says that generated revenues are spent for utility costs, medical supplies, for payroll and that a portion goes to pay for taxes. However, what it is really spent on which budget items nobody knows but administrators. We only see how much patients are paying and how much of that we do receive"*. Another physician from a different provider institution said: *"...I do not know where the rest of the money went, and if I decided to find this out I would be told by management to back off"*. Hospital Provider Gori. Our attempts to get for a sample of hospitals official budgets (which should have been public documents, hospitals being still under government ownership) were frustrated.

---

[279] Terminology used form Preker, AS, and A Harding, Eds., Innovations in Health Service Delivery: Volume I - The Corporatisation of Public Hospitals, Washington: World Bank, 2001.

8. 6.6 Impact of OOPs and of IPs

The study documented two important phenomena that result from the prevalence of OOPs and of IPs: 1) severe financial access barriers for the population; 2) the deteriorating condition of the health facilities.

1) Our study confirmed that households find it difficult to pay for health services, and often forgo, or delay or interrupt treatment because of financial reasons. Official figures show a dramatic decline in service utilization rates over the last decade. From 1990 to 1999 hospital admission rates plummeted by 68%, hospital *Bed Occupancy Rate* (BOR) decreased from 51% to 29% [GEPLAC-TACIS, 2001], utilization of outpatient services declined by 81% (from 7.3 visit per person per year to 1.42), and calls for ambulance services fell from 0.22 to 0.03 per person per year [CMSI, 1999]. Our study also found a significant difference between the burden posed by outpatient and inpatient services. Outpatient consultations, diagnostic tests and medications seemed to be relatively "affordable" for the majority of the population. Only the poorest (such as elderly or landless peasants) said they were forced to borrow from friends and relatives, or to sell valuables, to pay for outpatient services. For example, in a focus group, a representative from a poor household in Gori complained: *"...Two weeks ago I had poisoning and I needed 12 Gel for the medicines. My monthly salary is 15 Gel, and it is usually paid with significant delay. I had to borrow money and now I have to pay it back"*. By contrast, inpatient care cases, and particularly catastrophic ones (serious injuries and exacerbation of chronic diseases) are in general disruptive for households' finances for all socio-economic groups, with the exception of the very richest segments of the population. In cases of illnesses which require hospitalisation, the primary source of financing for patients is borrowing from relatives or friends. For example, a young man from a relatively well-off household in Tbilisi reported: *"My grandfather had urgent abdominal problems and we took him to the hospital. He had surgery, which cost approximately 1,100 Gel. This was quite significant for us and we had to seek support of friends in order to meet the costs"*. Other households said they had sold family assets and valuables to pay for health services. For example, a poor woman in Tbilisi reported to have sold her family's jewels to pay for health services: *"When I fell sick I was unemployed and had to borrow from relatives to cover the cost of treatment. Later on I sold some of my family's jewels to repay the debt."*

2) A second important result of IPs prevalence is that key inputs are severely under-funded. A couple of providers interviewed reported to use part of the revenue from IPs to improve quality of services (by purchasing extra-drugs, emergency tool-kits, etc.), but the majority admitted to pocket IPs as personal income. Important non-labour input components, particularly capital, are almost not funded at all, and as a result essential medical equipment or failing infrastructure cannot be adequately maintained or replaced[280].

8.7. Discussion and Conclusion

Figure 8.2 describes the *vicious cycle* in which the Georgia's health sector is trapped. In 1995, when the health reforms were launched, public financing for the sector was approximately US$ 0.8 per capita per annum, after the government's fiscal collapse in the first transition years (World Bank 1996). The Government of Georgia followed "the manual" in implementing Purchasing Reforms (see Chapter 6): it separated purchasing from provision; excluded health employees from the civil service, granted full-managerial autonomy to providers, moved from an input to an output-based payment system, and officially introduced user-charges. However, in the context of a catastrophic public health financing crisis and of extremely weak governance, such reforms de facto gave the government *imprimatur* to an unrestrained (wild?) market for health services that was at time taking shape in the country (Belli et al., 2002). Specifically, the new output-based payment system for providers contributed to generate an increasing reliance on formal and informal payments, by making health workers' income totally dependent on utilization of services. Health staff became progressively used and to some extent "forced" to ask patients to pay directly for services in order to make ends meet. In addition, the poor performance of public purchasers, the unclear definition of services in the BBP, and the widespread corruption among facility administrators and their 'regulators' created a situation where no accountability, no rules were left, and health care providers de facto became managed as private companies, with extremely high-powered economic incentives for their employees to attract patients, and to charge them as much as possible.

In addition, the rapidly declining GNP (Georgian "lost" approximately three quarters of its per capita real income during the first transition years.) and the dramatically increasing economic inequalities

---

[280] This progressive deterioration of the health system's infrastructure has been also documented elsewhere

severely limited patients' ability to pay and health providers' ability to raise revenue, particularly in rural areas. Hence, the emergence of OOPs as the main source of revenue for health care, under economic duress lead to a drastic decrease in demand, and further widened the gap between demand and supply, which was there under the communist regime. [281] The health sector became trapped in a plummeting spiral, where demand and supply are both progressively decreased, but where the gap between the two is never reabsorbed, described in Figure 8.2. As a result, presently in Georgia there is a vast area of unmet medical need, which is paradoxical given the country's excessive potential to supply health services.

Unless the above contextual or systemic factors are addressed, unless governance within the health sector and health financing of basic services are addressed, it will be difficult to contain informal payments. Other proposed changes may not render the expected results. For example, the policy recommendation to legalize user charges to avoid informal transactions would not lead to the desired results. On the supply side, it is important to strengthen the current process of rationalization, so far only timidly implemented. The elimination of existing duplications and of a significant number of under-utilised facilities and personnel will free some resources that can be used to improve quality of services in the facilities that the government decides to maintain. However, interventions on the supply-side are not sufficient. Priority should also be given to reverse the decline in demand, which can be done by increasing pre-paid resources for health, lowering prices at the point of service for essential and costly services such as hospitalization, and by empowering consumers through increased awareness about free entitlements in the BBP. The latter will also require revision of BBP to make it technically sound and easy understandable for the public.

[Zoizde et al. 1999].
[281] As discussed previously, the Georgian health system was characterized by excess capacity, both in terms of human as well as capital resources.

Figure 8. 2: The Vicious Cycle in the Georgian Health System

DEMAND

SUPPLY

CONTEXT

Decreased demand

Financial access barriers

Excess supply of providers relative to demand

Emergence of informal payments

Low income of providers

Overcapacity

Fiscal under performance and low level of priority assigned to health sector → low level of public funding

Poor performance of public purchasers

Unclear definition of the BBP

Lack of accountability in management of public facilities

Policy Change
1. Introduction of official OOP Payments
2. Purchaser-provider split
3. Autonomization of providers

Annex 8.1 Descriptions of Study Participants

Description of in-depth interviews participants (patients)

Description of households (HHs) that were selected for the in-depth interview is summarized in the following tables:

Table 8.3: Sample Matrix for In-depth Interviews with Patients

Location	Income Group	N	Percent
Tbilisi	Poor	21	36.2
	Rich	37	63.8
	Subtotal	58	100.0
Gori	Poor	13	86.7
	Rich	2	13.3
	Subtotal	15	100.0
Villages	Poor	22	84.6
	Rich	4	15.4
	Subtotal	26	100.0
Total	Poor	56	56.6
	Rich	43	43.4
Grand Total		99	100.0

Table 8.4: Average HH Monthly Total Expenditure per Geographic Location (Gel)

Location	Mean	95% Confidence Interval for Mean		Median	Minimum	Maximum
		Lower Bound	Upper Bound			
Tbilisi	422.3	337.0	507.6	400	50	2000
Gori	149.9	100.5	199.2	100	28	300
Villages	153.8	58.5	249.0	75	14	1200
Sample Average	310.5	249.2	371.8	200	14	2000

The total number of individuals residing in the HHs that were interviewed was equal to 406, out of which 146 individuals (36%) reported having used health services (because of illness, prevention or because they were pregnant) during the 3 month prior to interview. The average size of HHs differed by income groups and by type of residence, as follows:

Table 8.5: Average Size of HHs

	Poor	Rich	Residence			Total for Sample
			Tbilisi	Gori	Villages	
N	56	43	58	15	26	99
Mean size	3.8	4.6	3.8	4.6	4.6	4.1

Considering that only HHs that had used formal health services were selected for the interviews, 146 cases of service utilization were recorded. Details of service utilization are presented in the following tables:

Table 8.6: Type of Services Used by the Individuals from Selected HHs

Type of Service used	N	%
Outpatient polyclinic/ambulatory	35	24.0%
Specialist at hospital	39	26.7%
Private doctor	26	17.8%
Hospitalised	19	13.0%
Diagnostic Service	15	10.3%
Home visit by doctor/health professional	8	5.5%
Surgery	4	2.7%
Total	146	100.0

Table 8.7: Description of Cash Expenses by Different Income Groups (Gel)

	Poor		Rich	
	N. of cases	Mean payment	N. of cases	Mean payment
Total number of cases where cash payment occurred	58	186.7	65	543.0
Total number of cases	72	150.4	74	477.0

Table 8.8: Description of Expenses by Type of Residence (Gel)

	Tbilisi		Gori		Villages[282]	
	N	Mean	N	Mean	N	Mean
Total number of cases where cash payment occurred	80	450.3	16	119.3	27	303.2
Total	97	371.4	18	106.1	31	264.1

Table 8.9: Description of Expenses by Type of Services Used (Gel)

Type of service used	Number of cases	N. cases in which cash payment occurred	Drugs	Doctor's Fee	Hospital admission/stay	Surgery	Diagnostic service	In-kind supply (bed linen, food, etc.)
			Mean	Mean	Mean	Mean	Mean	Mean
Outpatient polyclinic/ambulatory	35	27	35.3	39.0	-	20.2	27.0	-
Specialist at hospital	39	34	75.9	21.4	31.0	36.0	31.8	1.5
Private doctor	26	24	40.9	17.7	0.2	-	47.7	-
Hospitalisation	23	18	154.0	46.9	75.7	234.1	40.4	18.2
Diagnostic Service	15	13	51.2	4.0	-	-	37.2	-
Home visit by doctor/health professional	8	7	147.1	30.4	-	-	3.1	-

---

[282] Includes transportations costs

Description of Focus Group Participants

A total of 119 individuals were selected to participate in the focus group discussions. Average age of participant was 39.2 years. The participant distribution by various age groups is presented in the following table:

Table 8.10: Focus Group Participant Distribution by Different Age Groups

	N	%
25-35 Years old	52	43.7
36-45 years old	32	26.9
46-55 years old	28	23.5
56-65 years old	5	4.2
66 + years old	2	1.7
Total	119	100.0

Participants were chosen according to the criteria discussed in the introduction:

Table 8.11: Participant Distribution by Various Groups and their Characteristics

Group	N	%	Mean HH Monthly Spending Gel
Tbilisi rich	29	24.4	634.8
Tbilisi poor	30	25.2	219.0
Gori rich	30	25.2	396.5
Gori poor	30	25.2	126.3
Total	119	100.0	344

Table 8.12: Educational Level of Participants by Income Groups %

	Tbilisi rich	Tbilisi poor	Gori rich	Gori poor
01 to 11 years of secondary school completed	10.3	43.3	23.3	36.7
Completed technical school (PTU or SPTU)	10.3	20.0	10.0	23.3
Completed higher/bachelors	79.3	33.3	60.0	40.0
Incomplete higher/bachelors		3.3	6.7	
Total	100.0	100.0	100.0	100.0
N	29	30	30	30

Table 8.13: Primary occupation of FG Participants by Income Groups %

	Tbilisi rich	Tbilisi poor	Gori rich	Gori poor
Unemployed	3.4	3.3	13.3	16.7
Self-employed		16.7	26.7	3.3
Business person/entrepreneur	13.8			
White collar worker	62.1	23.3	46.7	40.0
Agricultural worker			6.7	10.0
Retired/pensioner		33.3		
Domestic work	20.7	23.3	6.7	30.0
Total	100.0	100.0	100.0	100.0
N	29	30	30	30

Description of Providers

The study interviewed 30 individual providers, 20 physicians and 10 nurses. Among them, on average doctors had 18.2 years of experience and nurses 22.4.

Table 8.14: Distribution of Interviewed Doctors and Nurses by Specialty Area (%)

	Physician	Nurse
Obstetrics	20.0	20.0
Trauma	10.0	
Surgery	3.3	10.0
Department of general medicine	20.0	10.0
Laboratory	3.3	
Intensive care unit	3.3	
Department of the anaesthesiology	3.3	
Outpatient department		10.0
Insurance department	3.3	
Paediatric (children) department	10.0	20.0
Rehabilitation department	3.3	
Neurology department	6.7	
Cerebral palsy rehabilitation centre	3.3	

Diagnostic centre	3.3	
General medicine (adults)		10.0
Women's consultation	3.3	
Department infected pregnancy undelivered		10.0
Cardiology department	3.3	
Allergy department		10.0
Total	100.0	100.0
N	30	10

Most providers complained that their official salary was insufficient even for a normal existence, thus indicating a situation of profound unbalance in the system. Informal payments have been the way doctors and nurses have coped with the situation, transferring on patients most of the burden for financing government services.

Table 8.15: Sufficiency of Providers Income Perceived by Providers Themselves

	Physician	Nurse
Sufficient only for basic needs	6	1
Insufficient for a normal existence	24	9

Annex 8.2 Research Tools

Focus Group Guideline

Introduction

Moderator: Good evening and thank you for coming. We invited you to participate in this group discussion on the issues related to the costs of obtaining medical services in our country. This project is financed by the World Bank and it is aimed at improving the situation in our health care system, by developing with your assistance new policy recommendations. We are interested in patients' attitudes toward the following issues: how they evaluate recent health care reforms, what

experience they have in relation with different medical institutions, their experiences and perceptions on the costs of medical services. Our discussion will continue for 1 hour, and we will record it. In the final report we will describe your ideas only in general form, with confidentiality. In addition, we would like to stress that there is no right and wrong statements; we only want to hear your ideas about the issues.

Now please shortly introduce yourselves: name, age, profession and occupation.

Warm-UP

Let's discuss the current situation on health care in Georgia

- What do you think has changed in Georgia's healthcare during past several years?
- What are positive and negative new elements (strong and week points) these changes brought about?

Users' perceptions about medical care and payment

- Do you know of any free health services, financed by the government? Have you or any of your HH members and/or relative/friend recently received free services and if not, why?
- For those who received service for **free**, which medical services were provided for free and why? *(Probe: provider was friend; friend/relative paid for the services; public financiers paid the services; patient had private insurance that covered the cost of treatment).*
- In your experience, were services for children 0-3 less costly than the others?

Do you think that in general it has been difficult for you or your family to meet your health bills (ask about their family experience)?

- Do you see any relationship between quality and cost? If "yes", in which cases? Can you name specific examples?

If the government had enough money to provide just 3 services for free which ones do you think it should guarantee? Rank then (1, 2 and 3)

-Outpatient and long-term residential care for the elderly

-Emergency and hospitalisation services

-Surgery.

-Maternity and child services.

-Medicines

-Outpatient care for children under 15.

Why? *(Probe: Services who are best for improving health, services who are more costly, or people who deserve more)*

- Now, for all the other services, we must find other ways to fund them. What do you think it is the best way to pay for these other services?
- Do you prefer to pay when you get sick, or would you prefer to pay something each month in order to get services for free when you get sick?
- If you are willing to pay something each month, whom would you trust more for managing this money (pre-payment scheme, explain what we mean by it), so that when you get sick you are sure of getting (almost) free services:
- Central Government/MoLHSA
- Sakrebulo -Municipal government
- Polyclinic for Tbilisi, and health centre for rural areas
- Private insurance
- School
- Employer
- Labour union
- Other. Who?

Why?

- Which change would be needed to increase your trust? Can something be done to increase your trust in any of the above agents?
- Do you know how much different health services would cost you before you go? Do you think that having a price list posted at the door of any facility (polyclinic and hospital) would be helping you?
- If you were told that your health care expenses would be paid you back if you can present receipts, would you be able to demand and get receipts/formal confirmation from a doctor and/or a health care institution?

Attitude towards physicians, their motivation and regulation of payments

- How would you rate the size of a physician's income?

Privileged	1
Sufficient for a good standard of living	2
Sufficient only for basic needs	3
Insufficient for a normal existence	4

- How would you assess physicians' and nurses' income, in comparison to other professions? Are their income really so low? Please describe physicians and nurses separately.
- How would you assess incomes of people working in the area of the health administration?
- Do you think that informal payments/gifts increase substantially the incomes of physicians? Nurses? Health facility administrators? Please describe separately.
- What do you think, should the giving of presents or cash payment as an additional remuneration of physicians be regulated in some way, or banned, as a system? What is your recommendation, how should this be arranged?

Moderator: Thank you once more for coming.

Interview guide for medical providers

INFORMAL HEALTH PAYMENT IN GEORGIA

We are conducting a study financed by the World Bank and my name is

_____

The study concerns the Georgian health care system, and it is about the cost of obtaining medical services in our country. We are interested in provider' attitudes toward the following issues: how they evaluate recent health care reforms, what do they think about the existing economic situation of health care professionals, and, if there are informal payments in the system, why they exists, what impact they create on the profession and on the patients. If you know of such payments in your or in any other facility, we would also like to ask you how they are shared, and how they are used.

If you accept, we would like to interview you. Our discussion will continue approximately 45 minutes; and we will record it. In the final report, we will describe your ideas only in general form, without mentioning your name. We guarantee that the confidentiality of your statements will be always maintained. In addition, I would like to stress that there is no right and wrong statements. We only want to hear your honest views about the above issues.

Profile of respondent

1. Age |__|__|__| years
2. Sex   male   female
3. Years of experience |__|__|__| years
4. The specialty of the doctor/nurse   _____
5. The facility she/he works   _____
6. The department/ward she/he belongs _____
7. How many patients do you see every day? How many in total in the department where you work?
8. Do you think that in your department the numbers of doctors is adequate for the current level of activities/workload?
9. Does he/she have evening or night calls?

10. Can you please just describe your daily activities as a doctor/nurse (share of work in the ward, in the ambulatory within the hospital, outside, and at home)

Physicians' perceptions and experience

11. What are the daily difficulties in the physician's/nurse's work at your facility *(PROBE: work-load, quality and/or of medical supplies, conditions of the facility, the amount and the way they are paid, relationship with management, relationship with colleagues, relationship with patients etc...)?*

12. Has it changed since the health reforms started? In what way? Please provide details. *Please describe separately physicians and nurses*

13. How much and according to which criteria is she/he paid officially? *(e.g. fixed salary, capitation, fee for service, fee for procedure, mixed system)*

14. How would you rate the physicians/nurses' total income?

Physicians'		Nurses'	
Sufficient for a good standard of living	1	Sufficient for a good standard of living	1
Sufficient only for basic needs	2	Sufficient only for basic needs	2
Insufficient for a normal existence	3	Insufficient for a normal existence	3

15. What changes have occurred in your income since the reforms?

Financing of the facility/waiting lists

16. Do you think that your health facility has sufficient financial resources in order to provide the patients with the services they need?

Payments and Gifts

17. In your facility are there payments made by the patients? How frequently officially, and unofficially?

18.	Can you please describe in which form these payments occur *[PROBE: cash directly paid to doctor/nurse; there is always a person designated to collect all payments from patients; payment is made at the cashier's desk and there is a system to distribute the funds between providers, etc]*

19.	Can you please give us as many details as possible about the informal component of payments – *(Probe: ex-ante or ex-post, the form of payment gifts, in kind or money, whether patients in general accept to pay, how payment is determined, etc)*

20. What is the percentage of payments that goes officially to the facility? Give just your estimates; we are not looking for precise figures.

21.	For all the rest, who benefits? *[PROBE: the individual who receives; the money is distributed among various facility staff; to facility administrators; patient (please inquire why patient should benefit)*

22.	Who precisely administers these funds and how? When money is paid, is it distributed across a group of people, or it is paid individually to medical staff? If it is goes to team, is the distribution fair? Can you please describe the process in detail? *[PROBE: there is a set amount that every doctor/nurse should give to the facility, administration imposes the rules, collects the funds and makes decision; doctors/nurses accumulate agreed amount make joint decision how to spend it, etc]*

16. Does any portion of the amount paid by the patient go for drugs, or for the improvement of the services provided in your facility? If "yes" who decides how much should go for these purposes.

17. Can you explain how the rates for informal payment are set at your facility? *(Probe: Who sets rates? How rates are set? Is it negotiated on ad-hoc basis, or there are some established rates? Nobody sets the rates but patients pay as much as they can).*

18. Do you charge different prices for the same service to poor and rich? How do you discriminate rich and poor?

19. Do you prescribe more services from rich than from poor?

20. Would you say that payments are made voluntarily by patients or that providers demand these payments *[PROBE: check maybe doctors are not demanding, but nurses are, maybe administration demands, maybe there are other people demanding or other reasons for paying. Try to collect information on various individuals/departments in the facility]*

21. **If providers demand, then ask**: 'What are the reasons for their existence? *[PROBE: insufficient income for providers; widespread corruption; poor governmental policies; demands placed by facility administrators; short supply of equipment/material; high taxes.]*

22. Please prioritise them. Which one of the above reasons is most important?

23. **If patients pay voluntarily, then ask:** 'Why are patients accepting to pay voluntarily? *[Probe: they pay to skip queues, to receive more careful attention, to receive additional services, to express gratitude, to buy drugs or other materials/consumables while under treatment, etc...*

24. Do they think a treatment can be interrupted/stopped if a patient refuses or is unable to pay? Does this happen at your facility, and if yes how often?

Potential solutions, given the present financial constraint on the public side:

25. Do you think that informal payments are a serious problem? For patients? For doctors? Do you think the system as it is frustrates your expectations/ethical believes? Who really benefits from the existing situation?

So far, we have discussed how the situation is. Can we spend a few minutes discussing possible changes to the current system?

33. Do you think it would be advisable/desirable to legalise informal payments that occur at your facility? Please explain why you think so?

26. If all informal payments are legalized, what consequences they may have on providers and on patients? *[Probe: prices are more transparent, but health care might become more expensive].*

27. Do you think that the delivery system should be organized such that some services are provided for free and others are paid for? How should it work? For example, hospitals could develop private wards for paying patients alongside public wards for poor patients. Have you thought about this possibility? How do you think services in the two kinds of wards should be regulated?

28. Do you think it would be better to guarantee public insurance coverage for primary/outpatient care or for inpatient care, and for which reasons? Other possibilities?

29. Another possibility is to allow and encourage private practice outside the hospital, or inside, as intra-mural activity after regular working hours). Would it be easy for physicians/nurses to find patients outside the hospital? Is it a risky step?

Interview guide for users/patients

Interview Code _____

INTERVIEW GUIDE FOR USERS OF HEALTH SERVICES FOR THE INFORMAL PAYMENT FOR HEALTH IN GEORGIA

Interviewer : Good evening and thank you for coming. We invited you to participate in this group discussion on the issues related to the costs of obtaining medical services in our country. This project is financed by the World Bank, and it is aimed at improving the situation in our health care system, by developing with your assistance new policy recommendations. We are interested in patients' attitudes toward the following issues: how they evaluate recent health care reforms, what experience they have in relation with different medical institutions, their experiences and perceptions on the costs of medical services. Our discussion will continue for 1 hour, and we will record it. In the final report we will describe your ideas only in general form, with confidentiality. In addition, we would like to stress that there is no right and wrong statements; we only want to hear your ideas about the issues.

Now please shortly introduce yourselves: name, age, and occupation

1. Have you or any member of your household (HH, we will define as members of the same household people who live under the same roof and that share the same budget) used medical services during past 3 month?

   Yes   *Continue administering the questionnaire*

   No   *Move to different HH*

MODULE 1

2. For which particular health problem or any other reason (for example, pregnancy) did you use medical services?

3. Which medical services were used?

A. Outpatient (includes visit to polyclinic, ambulatory, health or, diagnostic service)	1	*Continue administering the questionnaire*
B. Hospitalised for more then 1 day	2	*Continue administering the questionnaire*
D. Alternative healthcare provider	4	*Move to Different HH*

Profile of respondent

4. Age   _|_|_|_| years

5. Sex   Male   Female

6. What is the highest level of education that you have attained? *Instruction: for those with only secondary school education, please indicate the number of years of education (from 01 to 11). For others use proposed coding.*

No education 00	Completed higher/bachelors 13
01 to 11 years of secondary	Incomplete higher/bachelors 14
school completed 01 to 11	Other 15
Completed technical school (PTU or SPTU) 12	Don't know/Refused to answer 99

7. What is your primary occupation? *Interviewer: if a respondent is engaged in several occupations, then ask "which one does the person spend the most time at" and report this as primary occupation.*

Pre-school 01	Unskilled worker 08
Pupil or Student 02	Agricultural worker 09
Unemployed 03	Retired/pensioner 10
Self-employed 04	Domestic work/homemaker 11
Business person/entrepreneur 05	Other 12
White collar worker 06	Don't know/refused to answer 99
Blue collar worker 07	

8. What sector is the person employed in? (Please ask this for only those respondents that are engaged in paid work)

Agriculture 01

Mining   02

Manufacturing   03

Electricity, gas, water   04

Construction 05

Trade 06

Hotels and Restaurants   07

Transportation and communication   08

Financial services and real estate   09

Public administration and defence   10

Education   11

Health and social services 12

Other personal services   13

9. How many people live in your household ? (Interviewer: if someone has been visiting, and staying with, the household for most of the last 3 months, then they should be included as household members)

⌐| | | |⌐

10. Approximately how much your HH spends on various things per month? Please come up with total figure and include all expenses that your HH faces (PROBE: food, utilities, transportation, education, leisure, healthcare, cigarettes and alcohol, presents etc)⌐| | | |⌐ lari

11. Please can you give us more details about the major episodes of utilization of health services (for simplicity referred to as "episodes of illness") in you household over the last three-month (complete table 1 below)

Table 1:   Major episodes of illness that occurred during last 3 month in the HH, followed by at least one doctor visit.

CODE		Name of person	Age	Health problem	Type of services used **	Total price paid
01	1st Episode of illness					
02	2nd episode					

03	3rd episode					
04	4th episode					

Instruction: Interviewer please fill-out the table for at most 4 episodes of illness, which required at least seeing a doctor. It is not mandatory to always collect 4 episodes.

** *Use following codes for the* <u>*Type of Service*</u>:

    a. Outpatient polyclinic/ambulatory,

    b. Specialist at hospital,

    c. Private doctor

    d. Hospitalised,

    e. Diagnostic

    f. Home visit by doctor/health professional

    g. Other _____(please specify)

Instructions for interviewer: MODULE 2 of the tool is administered separately for each episode of illness. For every new episode please take new set of the MODULE 2. Whenever you inquire about the episode please take a note of the name of the person from Table 1, the disease they had, and the services they used. Ask all the questions for this person and for the given episode. **Only and when you finish administering this MODULE 2, take new one and start with the next person and/or episode.**

<div align="center">

MODULE 2

Interview Code _____

Person/Disease Code from Table 1_____

</div>

Information, accessibility and affordability

    12. After the onset of illness did _____ *(the name of person from Table 1* immediately go to a provider or did s/he wait? If yes, why? *[PROBE: the illness was serious; had illness for a long time and knew how to treat it; had no money; provider is not available in close proximity, etc.]*

13. Can you please describe how s/he selected the provider? *(Probe: according to quality, proximity from residence, cost, through friendship, etc).* Please provide as much details as possible [in the case of outpatient, ask the individual who received care directly, if possible], or other member of household (HH)]

14. Have the criteria of selection changed during past years? If "yes" how and why?

15. What are the main difficulties your HH faced while had to go to a physician/hospital? [Probe: such as transportation costs, waiting lists, inability to pay for services, registration procedures, bad quality of available medical services, not having provider/hospital in close proximity.

Contact with clinical staff

16. Has he/she interrupted or not fulfilled completely treatment? How? [He/she was referred to a diagnostic service, but did not go, he/she did not buy all the drugs that were prescribed and he/she did not take all the drugs that he/she bought].

17. For what reason did he/she interrupt treatment? *[Probe: such as transportation costs, waiting lists, inability to pay for services, registration procedures, bad quality of available medical services, not having provider/hospital in close proximity].*

18. Whom did you go to receive treatment/care instead? [PROBE: drug shop, private physician, healer, friend/neighbour, other].

19. Has he/she aware of the possible consequences for her/his health of this interruption/ change of service?

Payments and gifts

20. Please give a description of expenses for that particular episode of illness *[PROBE: collect information separately for the payments made to facility; doctor; nurse; diagnostics; pharmaceuticals; other; specify which, private or public]*?

	Drugs	Physician consultation	Hospital admission/	Surgery	Diagnostic services	Nurse service	In kind supply of services	Other [283]	Total

---

[283] Includes everything else. Unfortunately, we did not single out a few other key cost items, such as transportations costs, and we are thus unable to precisely state what these 'other' expenses are. However, our impression is that transportation costs in fact was the largest item in the 'other' category.

			stay				(bed linen, food in hospital, etc)		
Episode of illness:									

21. Please give as many details as possible – the form of payment, for what services, when? To whom? Has he/she received any services for free? *[PROBE; was payment voluntary or demanded; was payment in cash or in-kind; were they asked to pay before (ex-ante) treatment or after (ex-post) treatment was completed etc]*

22. How did he/she receive information about payment and prices – *(ex. they know this from friends before visiting hospital/polyclinic, they saw the price list there, they sign the contract (?), the personal gave them information verbally, they saw price list).*

23. What was the consequence of payment on level/quality of service? Have he/she received better service? What would happen if he/she did not pay?

24. Did he/she recover fast? Was full recovery achieved?

25. Did he/she buy all the drugs that the doctor prescribed? If not, why?

26. During/after treatment, did he/she give gifts? If "Yes" did he/she give gifts besides payment or instead of them? To whom, in which form *(Probe: e.g. flowers, chocolates, cash or in another way?)*

27. How would they assess the fee for this episode of illness (unaffordable, very high, high, just, etc)?

28. What was the consequence of fee on family budget? *[Probe: they had to spend all cash income, all savings, take loan, and sell car, land, or house, borrowing from family and friends, borrowing elsewhere. If several sources were used, give precise share of each].*

29. Would you go to the same facility again? Why would you do so can you please explain?

MODULE 3

Interview Code _____

MODULE 3 OF THE TOOL IS ADMINISTERED FOR THE WHOLE HH AND NOT FOR THE INDIVIDUAL EPISODE OF ILLNESS

Willingness to pay for insurance, attitude towards government, perception of changes in the health system

30. Do you know any person in Georgia who received your same health services completely for free? *(Probe: provider was friend; friend/relative paid for the services; government paid the services; he had social or private insurance that covered the cost of treatment).*

31. Did you or any member of your household ever receive health services for free in the past, and if yes, why? *(Probe: provider was friend; friend/relative paid for the services; government paid the services; he had social or private insurance that covered the cost of treatment).*

32. Do you have children in your household under 3 years old?
   If no, skip next three questions:

   32.1 If yes, has he/she ever received medical services for free?

   32.2 Have you immunized him/her? Against which infection have you immunized?

   _____

   _____

   _____

   32.3 Have you paid anything to the health facility for this immunization?
   Yes   1   How much? _____
   No   2

33. What do you think has changed in the Georgian healthcare system during past few years?

34. What are advantages and disadvantages (strong and week points) these changes brought?

35. What sum would you willing to pay monthly for all member of your HH for the convenience of a physician to be at your disposal 24 hours per day and to receive services when you need them at no cost?

36. If you are willing to pay something each month for health insurance, whom would you trust more for managing this money, so that when you get sick you are sure of getting free services:

♦ Government: Central or Municipal

♦ Social insurance (SMIC)

♦ Private insurance

- Polyclinic for Tbilisi, and health centre for rural areas

- Private insurance

- School

- Employer

- Labour union.

- Other. Specify?

Why?

37. Which change would be needed to increase your trust? Can something be done to increase your trust in any of the above agents?

Attitude towards physicians, their motivation and regulation of payments

38. How would you rate the size of a physician's income?

Privileged	1
Sufficient for a good standard of living	2
Sufficient only for basic needs	3
Insufficient for a normal existence	4

39. How would you assess physicians' and nurses' income, in comparison to other professions? Are their income really so low? Please describe physicians and nurses separately.

40. How would you assess incomes of people working in the area of the health administration?

41. Do you think that informal payments/gifts increase substantially the incomes of physicians? Nurses? Health facility administrators? Please describe separately.

42. Should the giving of presents or cash payment as an additional remuneration of physicians be regulated in some way, or be banned? What is your recommendation, how should this be arranged?

Annex 8.3 The evolution of the health delivery system in Georgia

In this Annex we present some additional information concerning the evolution of the human resources employed in the health system over the last decade. Since initiation of health sector reforms in 1995, human resources have changed significantly. By 1999 the actual number of practicing physicians had decreased by almost 25% compared to 1991 levels, and the number of nurses by 59% *(See following figure).* However, due civil conflicts in some regions of Georgia, economic hardship that forced significant portion of population out of the country, declined birth rates, etc, which took place in the country during 1991-1999, total population sharply declined, so that physician to population ratio has not been affected significantly, hovering around 4.5 per 1,000 (this is one of the highest doctor to population ratios in the FSE; see *Error! Reference source not found. below.*

The situation with nurses proves to be very different, and characterized by a sharp decline of the nursing staff to population ratio (579 per 100,000 pop. in 1999 from 1058 in 1991). Presently, Georgia has one of the lowest nursing staff to population ratio, together with Armenia and Tajikistan (461 and 480 per 100,000 population respectively). Also the nurse to doctor ratio declined sharply and currently Georgia has the lowest ratio in the region -1.19 nurses per one doctor (compared to an average 1.84 for NIS) *(See Figure below)*

These changes resulted in a lower decline of nurses' income relative to doctors'. However, real value of annual officially reported income is still extremely low for both [284] (420-3,600 Gel for doctors and 360-720 Gel for nurse), and it hardly compares with the minimum individual annual subsistence level of 1,080 Gel (World Bank 2001).

---

[284] These figures are derived from our study and vary significantly by type of facility (e.g. general hospital, maternity home, polyclinic and also by geographic location).

Figure 8.3 Georgia's Health Human Resources circa 1991-1999
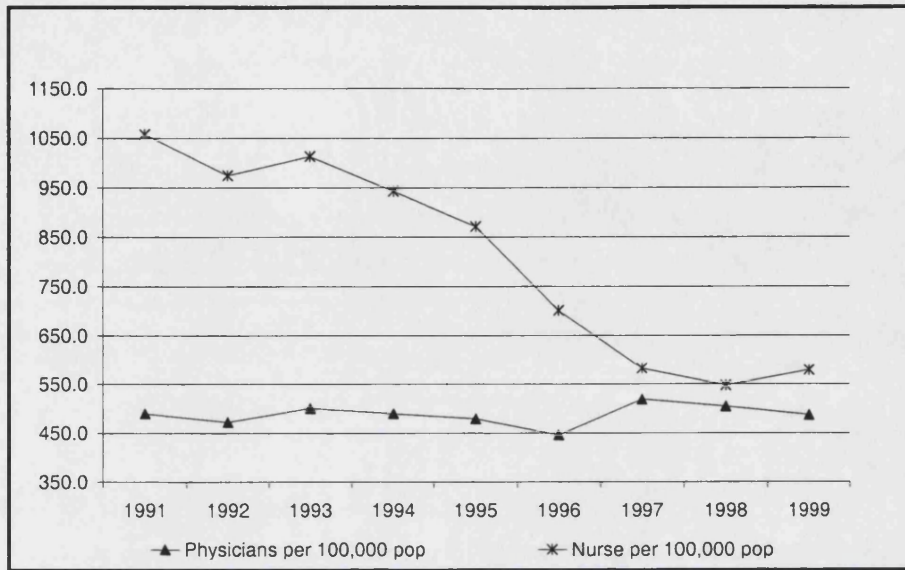
Source WHO Health for All Database

Figure 8.4 Nurses to Physician Ratio (1999 Data)
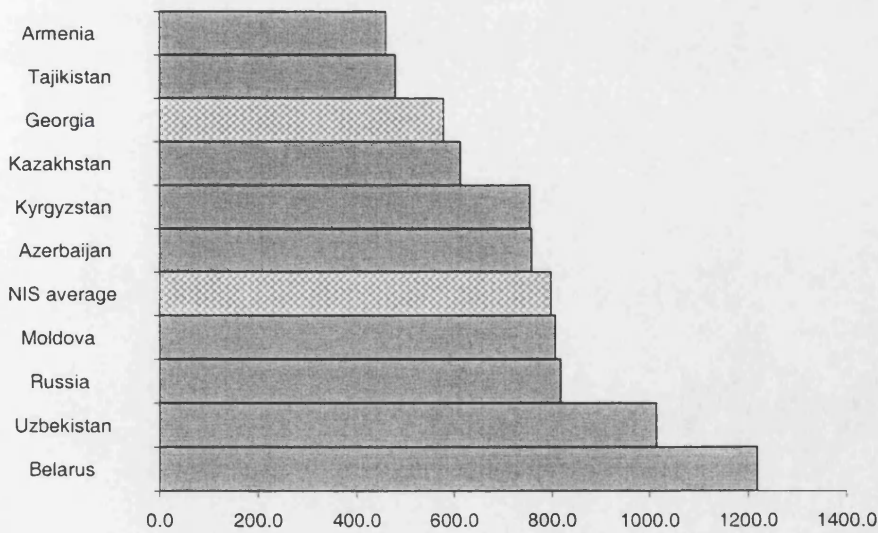
Source WHO Health for All Database

Table 8.16: Provider Network of Georgia, Circa 1988-1999

Type of Medical Facility (including non MoLHSA Facilities)	1988	1994	1995	1996	1997	1998	1999
Hospitals	402	321	289	264	287	272	287
Outpatient facilities in the hospitals	301	247	185	111	115	114	94
Independent outpatient facilities (ambulatories &polyclinics)	841	805	828	840	829	859	977
Dispensaries[285]	99	82	79	82	83	81	79
Independent dental clinics	56	53	53	51	51	56	62
Medical posts	613	502	180	33	47	53	33
Midwife posts	1,198	748	781	479	386	512	438
Independent ambulance stations	11	7	5	5	24	25	21
Blood banks	8	7	6	5	6	6	8
Infant orphanages	3	3	3	3	3	2	2
Scientific research institutes	19	19	19	19	19	19	18
Medical Centres						22	27	49

Source: *Zoidze et al. 1999*

---

[285] Dispensaries = Specialized outpatient facilities (e.g. TB Dispensaries, STD, Psycho-Neurological, etc.).

Annex 8.4: The Basic Benefit Package

The BBP design is very convoluted. It is hard even for healthcare professionals to understand what services are covered and what are not (Gotsadze and Rose 2000). Eligibility criteria for inclusion in the BBP are according to four criteria:

♦ special disease or condition (e.g. TB, Kidney failure, Oncology Disease, etc. under the State Medical Insurance Company, and a positive list of emergency conditions that cause catastrophic expenses under Municipal funding);

♦ age (e.g. children 0-3 years old);

♦ specific geographical location (e.g. residents of high mountainous areas, Tskhinvali residents);

♦ vulnerable status.

The historical evolution of the BBP also contributed to create its complex design. Initially only 6 programs were financed by SMIC portion of BBP (MoHLSA 1998). For the year 1999, the total number of programs included in the BBP increased to 14 (Schaapveld K. 2000). Partly, some high-value low-impact services were included (such as paediatric cardiac surgery, kidney transplantation, ischemic heart disease), and partly the BBP was expanded because some of the programs that initially were financed directly by the MoLHSA were included. With this increase, one trend became obvious – financial means did not meet ends. So, the balance was artificially achieved by lowering unit cost of reimbursement for services included in the BBP. All of this forced providers to seek resources directly from the population.

Services included in the BBP can be separated in three parts, based on the public purchasers that manage them (SMIC, PHD and Municipalities. See discussion in the text, Chapter 8, section on Health Reforms).

Table 8.17: Development of Central Government Funding and BBP During 1997-1999

1997	1998	1999
SMIC Programs	**SMIC Programs**	**SMIC Programs**
1.Treatment of Psychiatric Patients	1.Treatment of Psychiatric Patients	1.Treatment of Psychiatric Patients
2.Treatment of TB Patients	2.Treatment of TB Patients	2.Treatment of TB Patients
3.Prenatal Care and Delivery	3.Prenatal Care and Delivery	3.Prenatal Care and Delivery
4.Treatment of Children Under 1 years	4.Treatment of Children 0-2 years old	4.Treatment of Children 0-3 years old
5.Program for Vulnerable Population	5.Program for Vulnerable Population	5.Program for Vulnerable Population
6.Treatment of Oncology Patients	6.Treatment of Oncology Patients	6.Treatment of Oncology Patients
	7. Paediatric Cardio surgery	7.Infectious diseases
	8. Renal dialysis	8. Renal dialysis
	9. Infectious diseases	9. Surgical Treatment of Ischemic Heart Disease
		10.Organ transplantation
		11. Provision of Selected Chronic Patients with Pharmaceuticals
		12. Rural PHC Program
		13. Special medical program for Tskhinvali*
		14. Additional Medical Care for Highlanders**
MoHLSA Programs	**MoHLSA Programs**	**MoHLSA Programs**
1.Additional Medical Care for Highlanders	1.Additional Medical Care for Highlanders	
2.Renal Dialysis	2.Program for State border regions	
3.Diabetis Mellitus Program	3.Diabetis Mellitus Program	
4.Additional Medical Care for War Veterans	4.Program for special patients	
5.Medical program for mandatory military service	5.Medical program for mandatory military service	
6.Medical program for orphans in orphanages	6.Medical program for orphans in orphanages	
7. Paediatric cardio surgery program	7. Surgical Treatment of Alchemic Heart Disease	
8.Medical Emergency Programs	8.Medical Emergency Programs	1.Medical program for mandatory military service
9.Medical science and education	9.Medical science and education	
10.Critical medical facility	10.Critical medical facility	2.Medical Emergency Programs

1997	1998	1999
rehabilitation program	rehabilitation program	3.Medical science and education
11.Management of reforms	11.Management of reforms	4.Critical medical facility rehabilitation program
12.State sanitary surveillance	12.State sanitary surveillance	5.Management of reforms
	13.Medical criminal expertise	6.State sanitary surveillance
		7. Medical criminal expertise
PHD Programs	*PHD Programs*	*PHD Programs*
1.Expanded Program of Immunization	1.Expanded Program of Immunization	1.Expanded Program of Immunization
2.Prevention of Infectious Diseases	2.Prevention of Infectious Diseases	2.Prevention of Infectious Diseases
3.Health Promotion	3.Health Promotion	3.Health Promotion
4.STD Program	4.STD Program	4.STD, AIDS & safe blood program
5.AIDS Program	5.AIDS Program	5.Medical Information program
6.Medical Information program	6.Medical Information program	6.Prevention of drug addiction
7.Prevention of drug addiction	7.Prevention of drug addiction	7.Radiation and Iodine Deficiency
8.Radiation and Iodine Deficiency	8.Radiation and Iodine Deficiency	8. Active screening of diseases
9.Safe blood program	9.Safe blood program	9. Micronutrient Deficiency
10. Active screening of diseases	10. Active screening of diseases	10. Screening of cancer diseases
	11. Micronutrient deficiency program	11. Screening of cardiovascular diseases
		12. Trauma prevention

Chapter 9: Conclusions

The thesis has analysed the resource allocation and purchasing reforms implemented over the last two decades in several high, middle and lower income countries from an economic perspective. The premise has been the investigation of the equilibrium achieved by markets in health financing and health provision without government intervention, described in the first three chapters. These chapters have also anticipated some of the later results, by showing the impact of a plurality of possible mechanisms, including quality regulation, taxes and subsidies to private provision, which the government can implement to correct market outcomes.

In its central part, the thesis has focused on the impact of the new incentive system introduced with the RAP reforms on efficiency and on equity. Our analysis of the impact of health reforms has two important limitations. One is due to lack of rigorous, experimental evidence on the impact of reforms: with few exceptions, it seems that an extremely limited amount of resources have been devoted to monitor and evaluate ongoing reform efforts, even in high income countries such as West European countries. The quantitative evidence available is generally limited to a few aggregated indicators, such as number of beds, or number of physicians, or number of inpatient and outpatient cases, or average length of stay in hospitals, on the basis of which it is very difficult to ascertain whether health reforms achieved their intended results. This essay is very much constrained by such dearth of rigorous micro-level experimental data, and several of its conclusions are based more on direct experience, qualitative investigation, policy documents or on theoretical speculation rather than on conclusive empirical evidence.

The second limitation is self-imposed, and it stems from the thesis's attempt to draw some conclusions of general validity from the experience with health reforms. Focusing on the unifying features across so many and so diverse regions and countries, the thesis cannot and does not fully take into account all the important individual differences, and it does not investigate in detail single reform components in individual countries.[286]

---

[286] An introduction to single reform components in individual countries is in the Case Studies presented in each of the chapters.

In an attempt to gain in specificity, we have limited our attention to the resource allocation, purchasing and payment systems dimensions of the reforms. These reforms are a means to respond to the combination of market and government failures which prevent health systems from achieving better performance in terms of efficiency equity and quality of services. In Chapter 4 we discussed in detail the efficiency and quality and in Chapter 6 the equity deficiencies that affect health systems. We also described the main components of RAP reforms. Our interpretation is that RAP reforms lead to the emergence of trade-offs among potentially equally desirable aims, such as consumer responsiveness, cost consciousness, efficiency and equity. We have shown how each RAP reform component (for example, a particular payment system as described in Chapter 5) is in general appropriate to achieve one of those aims, and, by doing so, it moves the health system further away from achieving one or more of the other ones. Given the existence of trade-offs, we indicated that "optimal" health reforms should find a second-best equilibrium –or a balance- between the above trade-offs, and should find within the reforms' tool-kit a mix of interventions able to maximize positive impact and to control potential adverse effects.

In Chapter 6 we showed that the existing allocations of inputs and practices within the public sector, which have disproportionably benefited the rich in low and middle income countries, have proved exceedingly difficult to change. The purchaser-provider split, the new resource allocation and payment systems per se are insufficient to induce the rapid change that is expected. However, we also reviewed evidence showing that the new allocations mechanisms and incentives characterized by more transparent geographical distribution of resources, output/outcome orientation, arms-length relationship with providers, etc. have in several cases lead to real improvements in resource allocation, better quality of services and greater coverage of the poor. Their principal merit has been to increase the focus on results and performance (mainly in terms of outputs), away from passive, input based allocation mechanisms, and thus to increase the "value for money" for any given level of spending. At the same time, their full potential is not yet exploited. We have also indicated that successfully implementing RAP reforms requires governments able to exercise their steering/oversight role, which in turn presupposes a health financing framework and other preconditions which may not be yet feasible in the poorest, capacity constrained countries.

When adequately implemented, RAP reforms have also been able –through the purchasing/contracting mechanisms- to make a better use of private providers and to increase coverage of privately provided essential services, also to the benefit of the poor (see Chapter 6, analysis of the contracting out and in experiences). The private sector is potentially able to bring

about a more efficient and consumer responsive delivery system, provided the government can put in place a credible and enforceable regulatory framework. Enhancing quality of care, through MQS (see Chapter 3), containing phenomena of supply-induced demand (Chapter 5), and providing incentives (financial and others) to expand coverage to the poor and all those unable to pay on an OPP basis, should be priority goals in regulation of the private sector.

Finally, the analysis of the last two chapters indicates that the link of RAP mechanisms with the sources of revenue and pooling arrangements[287] is crucial: RAP reforms change the criteria according to which health resources are allocated and services purchased and paid, but they are embedded within a specific revenue collection or resource mobilization system, which determines the limits of what any RAP mechanism can achieve. The fundamental limitation of the reform process in several former socialist economies (FSE) countries, including Georgia, has been precisely the lack of a clear health financing strategy, to change the organization and the share of the different sources of revenue for health. Several sources of revenue collection have continued to be in place, each of them responsible for financing different services for different population subgroups: (a) contributions to a social insurance fund, for those in formal employment; (b) taxes and direct transfers pooled by Ministry of Finance and channelled through Ministry of Health, supposedly for the poor and destitute, in reality spread out to finance a plurality of programs and allocated according to arbitrary mechanisms; (c) direct transfers to public providers by local governments, to subsidize specific services, such as emergency services in Georgia; (d) private insurance contributions, for the rich, and (d) OOPs for all those non formally employed, such as rural workers. In fact, during the decade of the '90s, OOP payments have become the prevalent source of revenue, as all have been asked to pay directly for health services in the countries which were hit the hardest during the transition years.

Health financing reforms are to confront the scarcity of resources and not only inefficiency in resource allocation and use, which would include the mobilization and the pooling elements (see Figure 4.1 and Gilson et al. 2000). As we saw in Chapter 2, governments have a key role to play in

---

[287] The principal revenue collection systems are: (a) Out-of-pocket payments; (b) Voluntary insurance contributions (usually risk-related) with competing insurers; (c) Tax-based public funding; (d) Mandated employment-based insurance. The prevalence in the funding of any health system of one of the above funding systems determines the overall efficiency/equity impact in the distribution of the health financial burden. International evidence from industrialized countries shows that the only progressive funding mechanism is direct taxation. Employment-based social insurance contributions are generally proportional or mildly regressive, whilst private insurance contributions tend to be regressive (Van Doorslaer, 1997).

addressing these issues of resource mobilization[288], to guarantee an equitable risk sharing across individuals. By contrast, as the evidence presented in Chapter 8 demonstrates, where out-of-pocket payments are the prevailing source of revenue for health services, and/or health funding system is extremely fragmented, RAP mechanisms are largely irrelevant, and RAP reforms are doomed to fail. Market failures similar to those analysed in the first three chapters, namely exclusion of patients from insurance coverage and essential services, market segmentation and low quality of care for the poor, supply induced demand, cannot be eliminated.

Also within a funding system exclusively based on voluntary insurance premiums, cross-subsidies across risk/income groups are extremely hard to maintain due to risk-selection and adverse selection problems (Chapter 2). The only way the government can pursue equity-motivated corrections to market outcomes is through direct subsidies to the poorer and/or higher risk individuals, or to health plans that accept to offer coverage to the poorer, and higher risks. All pooling schemes based on individual voluntary participation will always be highly constrained in the amount of redistribution they can achieve precisely because they demand universal consensus of participants.[289]

The experience with health reforms in FSE also indicates that the new organizational model proposed for health, known as "managed competition" or "internal markets", is perhaps too fragile and institutionally demanding for countries with poor legal and administrative underpinnings, or weak governance. Only in the few countries where such pre-conditions were at least partially in place, such as some Central and East European countries and Colombia, have health reforms' results been more promising.

---

[288] Resource mobilization addresses questions regarding the introduction or removal of user fees; restructuring fees; the development of exemption policies; the development and introduction of insurance schemes, such as social, community-based, or private insurance; and the introduction of prepayment schemes.

[289] In the language of welfare economics, they will promote only cross-subsidies that lead to Pareto-improvements.

References

Abel-Smith, B., Falkingham, J., 1996, *Financing Health Services in Krygyzstan: The Extent of Private Payments*. Report for the British Government Know-How Fund (KHF). London School of Economics, London, U.K.

Akin, J, Hutchinson, P., 1999, Health care facility choice and the phenomenon of bypassing, *Health Policy and Planning* 14: 135-51.

Alderman, H., Lavy, V., 1996, Household Responses to Public Health Services: Cost and Quality Trade-offs, *World Bank Research Observer* 11: 3-22. World Bank, Washington.

Allen, R., Gertler, P., 1991, Regulation and the Provision of Quality to Heterogeneous Consumers: the Case of Prospective Pricing of Medical Services, *Journal of Regulatory Economics 3*: 361-375

Anand, S., Ravaillon, M., 1993, Human development in poor countries: on the role of private incomes and public services. *Journal of Economic Perspectives* 7(1): 133-150.

Anand, S., Sen, A., 1997, Concepts of human development and poverty: a multidimensional perspective', in *Human Development Papers, Poverty and Human Development*. New York: HDRO, The United Nations Development Programme, pp: 1-19.

Armstrong, M., Cowan S. and J. Vickers, 1994, *Regulatory Reform*, London, MIT Press.

Arrow, K., Debreu, G., 1954, Existence of an Equilibrium for a Competitive Economy., *Econometrica 22*: 265-290.

Arrow, K. Uncertainty and the welfare economics of medical care American Economic Review 1963; 53: 941-973.

Atkinson, A., 1970, On the Measurement of Inequality, *Journal of Economic Theory*, pp. 244-63.

Atkinson, A., Stiglitz, J., 1980, *Lectures in Public Economics*, New York. McGraw-Hill.

Audit Commission, 1997, *What the Doctor Ordered*, London, HMSO.

Auriol, E., Laffont, J., 1993, Regulation by Duopoly, *Journal of Economics and Management Strategy 3*

Baeza, C., Cabezas, M., 1999, Is there a Need for Risk Adjustment in Competing Health Insurance Markets in Latin America? Santiago, CLAISS

Balabanova, D., McKee, M. 2000, Undertsanding Informal Payments for Health Care: the Example of Bulgaria. Health Policy 2002; 62(3): 243-73.

Bangladesh Public Expenditure Review 1999/2000, Health Economics Unit, Ministry of Health and Family Welfare of Bangladesh.

Baron, D., Besanko, D., 1987, Monitoring, moral hazard, asymmetric information, and risk-sharing in procurement contracting, *Rand Journal of Economics 18*: 509-532.

Baron, D., Myerson, R., 1982, Regulating a monopolist with unknown Costs, *Econometrica 50/4*.

Barr, N., 1987, *The Economics of the Welfare State*. London Weidenfeld and Nicolson.

Beaglehole R, Bonita R., 1997, *Public Health at the Crossroads*. Cambridge: Cambridge University Press.

Beaglehole, R., Bonita, R., 1998, Public health at the crossroads: which way forward? *Lancet* 351: 590-3.

Beesley, M., Littlechild, S., 1989, The Regulation of Privatised Monopolies in the UK, *RAND Journal of Economics 98*: 371-400.

Beith, A., Eichler, R., Sanderson, J. and D. Weil, 2001, *Can Incentives and Enablers Improve the Performance of Tuberculosis Control Programs? Analytical Framework, Catalogue of Experiences and Literature Review*, Paper prepared for the Workshop on Incentives and Enablers To Improve Performance of TB Control, Paris, November 2001.

Belli, P., 1993, *Redistributive Effects of the Italian Health Service*, in: Paci, M. (ed.), *C.E.S.P.E. Foundation Report on Social Inequality in Italy*, Bologna, il Mulino.

Belli, P., 1997, *The Theory of Property Rights and the Purchaser-Provider Split in the Health Sector*, in: Petretto, A. (eds.), *The Reform of the Italian Health Service*, Bologna, il Mulino.

Belli, P., 2001a, How Adverse Selection Affects the Health Insurance Market, *World Bank Policy Research Working Paper*, No. 2148. World Bank, Washington.

Belli, P., 2001b, *Ten years of health reforms in the ECA region. Lessons learned and options for the future*, WP, Center for Population and Development Studies, V.11 n.6, Cambridge. MA.

Belli , P. Berman, P., Bossert, T., 2002, *Formal and Informal Household Spending on Health: a Multi country Study in Central and Eastern Europe*. Harvard School of Public Health, International Health Systems Group, 2002 (unpublished).

Belli, P., Gotsaddze, G., Shahriari, H., 2004, Out-of-pocket Payments in the Health Sector: evidence from Georgia, *Health Policy 70*, 109-123

Bennett, S., A. Creese, and R. Monasch, 1998, *Health insurance schemes for people outside formal sector employment*, World Health Organisation.

Bennett, S., McPake, B., Mills, A. (eds.), 1997, *Private Health Providers in Developing Countries*. Serving the Public Interest? Zed Press, London, UK.

Bennett.S and Gotsadze.G., 2001, *Rebuilding Risk Pooling Schemes in Georgia: Issues of Trust and Accountability*. Paper presented at the iHEA 3rd International Conference materials, July 22-25, York, UK.

Berman, P., Bossert, T., 2000, *A Decade of health sector reform in developing countries: what have we learned?* DDM Symposium, Appraising a decade of health sector reforms in developing countries, March 15th Washington DC.

Bevan, G., Bartlett, W. and S. Murphy, 1995, Transactors, Property Rights and Competition in the NHS Market, presented at the *Third European Conference on Health Economics*, Stockholm, Sweden.

Bidani, B. and Ravallion, M. 1997. Decomposing social indicators using distributional data. *Journal of Econometrics* 77: 125-139.

Birdsall, N. and Hecht, R., 1995, Swimming against the tide: Strategies for Improving Equity in Health. *HRO Working Papers*, the World Bank.

Bitran, R., 1998, *Equity in the Financing of Social Security in Chile*, Partnerships for Health Reforms, Abt

Associates, Bethesda, Maryland.

Black, D., 1980, *Inequalities in Health*, London, HMSO.

Blomqvist, A., 1991, The Doctor as Double Agent: Information Asymmetry, Health Insurance, and Medical Care, *Journal of Health Economics 10*: 411-432.

Bloom, D., Cunning, D., 2001, *The Health and Poverty of Nations, From Theory to Practice*. Paper for WHO Commission on Macroeconomics and Health.

Bloom, D., Malaney, P., 1998, Macroeconomic Consequences of the Russian Mortality Crisis, *World Development 26*: 2073-2085.

Bobadilla, L., Costello, C., Mitchell, F., (eds.), 1997, *Premature Death in the New Independent States*. Washington D.C.: National Academy Press.

Bobak, M., Marmot, M., 1996, East-west mortality divide and its potential explanation: proposed research agenda. *British Medical Journal 312*:421-425.

Bonilla-Chachin, M, Hammer, J., 2000, *Life and Death Among The Poorest*. Draft, The World Bank, Washington DC.

Bossert, T., 1993, *Lesson from the Chilean Model of Decentralization: Devolution of Primary Care to Municipal Authorities*. LAC Health and Nutrition Sustainability, Washington, DC.

Bossert, T., 2000a, Analyzing the decentralization of health systems in developing countries: Decision space, innovations and performance, *Social Science and Medicine*, 47: 1513-27

Bossert T., 2000b. *Decentralization of Health Systems: Chile, Colombia and Bolivia: Latin American and Caribbean Regional Health Sector Reform Initiative*, Report 29.
http://www.hsph.harvard.edu/ihsg/publications/pdf/lac/decentralization45.pdf

Brennan, G., and Buchanan, J., 1977, Towards a Tax Constitution for Leviathan, *Journal of Public Economics 8*: 255-74.

Brijlal V., Gilson L., Makan B., McIntyre D., 1997, *District financing in Support of Equity*. Johannesburg, Centre for Health Policy, Community Health, University of Witwatersrand.

Brown, L., 1991, *The National Politics of Oregon's Rationing Plan*, Health Affairs, Summer.

Brown, W., Churchill, C., 1999, *Micro-enterprise Best Practices Providing insurance to low-income households. Part 1: A primer on insurance principles and products.*
http://www.mip.org/pubs/mbp/providing_insurance_to_low_income_households-part1.htm

Brugha, R. and Zwi, A., 1998, Improving quality of privately provided public health care in low and middle income countries: challenges and strategies. *Health Policy and Planning 13(2)*: 107:120.

Buchanan, J., Tullock, G., 1962, *The Calculus of Consent*. University of Michigan Press.

Carr, D., Gwatkin D., Fragueiro D., and Pande, R..,1999, *A guide to country-level information about equity, poverty, and health available from multi-country research programs*. World Bank, Washington DC.

Carr-Hill, R., Hardman, G., Martin, S., Peacock, S., Sheldon, T. and Smith, P., 1994, *A Formulae for Distributing NHS Revenue Based on Small Area Use of Hospital Beds*. Centre for Health Economics, Occasional Papers.

Castro-Leal, F., Dayton J., Demery, L. and K. Mehra, 1997, Public social spending in Africa: do the poor benefit? *The World Bank Research Observer* (December). Washington DC.

Chakraborty, S., Harding, A., 2001, *The Private Health Sector Handbook, Chapter 2: Conducting a Private Health Sector Assessment (PHSA)*, The World Bank, Washington, DC.

Chalkley,M., Malcomson J., 1995, Contracting for health services with unmonitored quality, *Discussion Papers in Economics and Econometrics 9510*, University of Southampton

Chalkley M., Malcomson, J., 1996, Contracts for the National Health Service, *The Economic Journal 106*: 1691-1702.

Chalkley, M., Malcomson, J., 1998, Contracting for Health Services when Patient Demand Does not Reflect Quality, *Journal of Health Economics 17*: 1-20.

Chawla, M. et al., 1999, Enrolment procedures and self-selection of patients: evidence from a Polish family practice, *Health Policy and Planning 14*: 285-290.

Chawla, M. 2001, *How Well does the health sector serve the poor in Georgia: An Examination of Public and Out-of-Pocket Expenditures on Health (Draft)*, The World Bank, Washington, DC.

Chawla, M., Berman, P., and D. Dudarewicz, 1998, *Innovations in provider payment systems in transitional economies: experience in Suwalki, Poland*, mimeo

Chawla, M., Berman, P., Windak, A., Kulis M., 1999, *Provision of Ambulatory Health Services in Poland: a Case Study from Krakow*. International Health Systems Group, Harvard School of Public Health, Boston, Mass.

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

Chellaraj, C., Heleniak, T. and Staines, V., 1999, *Health Sector Statistics for Former Soviet Union*. World Bank Technical Paper. Social Challenges of Transition Series. Draft

Chenet, L. et al., *Alcohol and Cardiovascular Mortality in Moscow: New Evidence of a Causal Association*, Journal of Epidemiology and Community Health 52:772-774.

Choi, C and H. Shin, 1992, A Comment on a Model of Vertical Product Differentiation, Journal of Industrial Economics 60:229-231.

CMSI (Center for Medical Statistics and Information), 1999, *Healthcare in Georgia, 1998*: Statistical Yearbook. Tbilisi.

CMSI, 2000, *Healthcare in Georgia, 1999*: Statistical Yearbook. Tbilisi.

Cornia, A. and R. Panicchia, 1995, *The demographic impact of sudden impoverishment: Eastern Europe during the 1989-1994 transition*. In: Unicef Innocenti Occasional Papers, Ec. Pol. St. 49.

Crampes, C and A. Hollander, 1995, Duopoly and Quality Standards, *European Ec. Review 39*:71-82.

Crane, T., 1992, The Problem of Physician Self-referral Under the Medicare and Medicaid Anti-kickback statute, *Journal of the American Medical Association 268*: 85-91.

Cremer, H., Marchand, M. and J. Thisse, 1989, The Public Firm as an Instrument for Regulating an Oligopolistic Market, *Oxford Ec. Papers 41*: 283-301.

Crocker, K. and A. Snow, 1985, The Efficiency of Competitive Equilibria in Insurance Markets with Asymmetric Information, *Journal of Public Economics 26*, 207-219

Cromwell, J., Mitchell, J. , 1986, Physician Induced Demand for Surgery, *Journal of Health Economics 5*: 293-313

Culyer, A., Van Doorslaer, E., and Wagstaff, A, 1992, Comment: Utilization as a Measure of Equity, *Journal of Health Economics, 11*: 93-98.

Culyer, A., Wagstaff, A 1993 Equity and Inequality in health and health care. *Journal of Health Economics* 11: 207-210.

Curatio Intern. Foundation, 1997, *Hospital Financing Study for Georgia.*

Cutler, D., 1996, Public Policy for Health Care, *NBER Working Paper 5591*, May

Cutler, D., and R. Zeckhauser, 1997, Adverse Selection in Health Insurance, *NBER Working Paper n. 6107*

Dahlby, B. , 1981, Adverse Selection and Pareto Improvements Through Compulsory Insurance, *Public Choice 37*, 547-558

Daly, M.., Duncan, G., Kaplan, G , Lynch, J. , 1998, Macro to Micro Links in the Relation Between Income Inequality and Mortality. *Millbank Quarterly* 76, 315-339.

Daniels, N., Kennedy, B. P., Kawachi, I., 1999. Why justice is good for our health: the social determinants of health inequalities. *Daedalus* 128 (4): 215-251.

Davey Smith, G., Neaton, J., Wentworth, D.., Stamler, R., Stamler, J., 1998, Mortality differences between black and white men in the USA: contribution of income and other risk factors among men screened for the MRFIT. *Lancet 351 (*9107): 934-6.

De Fraja, G, 2000, Contracts for Health Care and asymmetric information, *Journal of Health Economics, 19*: 663-677.

De Fraja, G, and F. Delbono, 1989,Alternative Strategies of a Public Enterprise in Oligopoly, *Oxford Ec. Papers 41*:302-311.

Delbono, F., Denicolo', V. and C. Scarpa, 1996, Quality Choice in a Vertically Differentiated Mixed Duopoly, *Economic Notes by Banca del Monte dei Paschi di Siena 25*:33-46.

Delcheva, E., D. Balabanova, and M. McKee. 1997, Under-the-Counter Payments for Health Care. *Health Policy*, 42:89–100.

Denzin, N.K, Lincoln, Y.S. (editors), 2000, Handbook of Qualitative Research, 2nd edition, Sage Publications Inc.

Department of Health, 1989,*Working for Patients*, Cm 855, HMSO, London

Desai, M (1990). Poverty and capability: towards an empirically implementable measure (London: The *Development Economics Research Programme # 27*, STICERD, LSE).

DFID Health Systems Resource Centre. 2000. *Making the most of the private sector.* A workshop organised on behalf of the UK Department for International Development, May 11 and 12, 2000. London: DFID Health Systems Resource Centre.

Diamond, P., 1992, Organizing the Health Insurance Market, *Econometrica 60*: 1233-54

Diamond, P., and Mirrlees, J., 1971, Optimal Taxation and Public Production: optimal prodution and efficiency and II tax rules, *American Economic Review* 61

Diop, F., Yazbeck, A. and Bitran, R., 1995, The Impact of alternative cost recovery schemes on access and equity in Niger, *Health Policy and Planning*, 10(3):223-240.

Dixon, J., and H. Welch, 1991, Priority Setting: Lessons from Oregon, *The Lancet* vol. 337.

Doebbler, C., 2001 The right to health of children and the World Bank, Health and *Human Rights* 5: 3-29.

Doherty J, van den Heever A, 1997. A Resource Allocation Formula in Support of Equity and Primary Health Care. Center for Health Policy, Department of Community Health , University of Witwatersrand

Doherty J, van den Heever A, 1997. Re-distributing Health Care Resources on a Geographical Basis: Lessons Learned in South Africa, Centre for Health Policy , Community Health University of Witwatersrand

Doll R., Peto R.et al., 1993, *Tobacco and death in Eastern Europe*. In: Bodmer W., Zaridze, eds., Cancer Prevention in Europe. Organization of European Cancer Inst., Lodon:71-97

Dorabawila, V., Lewis, M. and V. Staines, 1999. *Health Status and Private Health Expenditures in the Kyrgyz Republic, 1993 to 1997.* World Bank, Washington, D.C. Processed.

Dranove, D., 1987, Rate Setting By Diagnostic Related Groups and Hospital Specialization, *RAND Journal of Economics 18*: 417-427

Dranove, D., 1988, Demand Inducement and the Physician/Patient Relationship, *Economic Inquiry 26*: 281-298

Dranove, D., Satterthwaite, M., 1992, Monopolistic Competition When Price and Quality Are not Perfectly Observable, *RAND Journal of Economics 23*: 518-534

Dranove, D., Shanley, M., White, W., 1993, Price and Concentration in Local Hospital Markets: the Switch from Patient-Driven to Payer-Driven Competition, *Journal of Law and Economics 36*:179-204.

Dranove, D., Wehner, P., 1994, Physician-Induced Demand for Child Births. *Journal of Health Economics 13*: 61-73.

Dranove, D., White, W., 1994, Recent Theory and Evidence on Competition in Hospital Markets, *Journal of Economics and Management Strategy 3*: 169-209.

Eberstadt, N., 1999, Russia: Too Sick to Matter? *Policy Review* 95: 3-27.

Ecchia, G., Albertini, L., 1997, Minimum Quality Standards and Collusion, *Journal of Industrial Economics XLV*:101-113..

Ecob, R., Davey-Smith, G., 1999. Income and health: what is the nature of the relationship. *Social Science and Medicine* 48, 693-705.

Ellis, R., Mc Guire, T., 1993, Supply-side and demand-side cost sharing in health care, *Journal of Economic Perspectives 7*, 135-151.

Ellis, R. and McGuire, T. Optimal payment systems for health services. *Journal of Health Economics* 1990; 9: 375-396.

Ellis, R., Mc Guire, T., 1986, Provider Behaviour Under Prospective Reimbursement, *Journal of Health*

Economics 5: 129-151

Ellis, R., Mc Guire, T., 1990, Optimal Payment Systems for Health Services, *Journal of Health Economics 9*: 375-396

Encinosa, W., Sappington, D., 1997, Competition Among Health Maintenance Organizations, *Journal of Economics and Management Strategy 6*, 129-150.

Ensor, T., Langenbrunner, J., 2000, *Allocating resources and paying providers*, World Bank (Draft).

Ensor, T., Killingsworth, J. (Editors), 2000, *Paying for Health: unofficial payments for Health Care in Low and Middle Income Countries.Cheltenham: Edward Elgar.*

Ensor, T. and L. Savelyeva, 1998, Informal payments for health care in the former Soviet Union: some evidence from Kazakhstan, *Health Policy and Planning* 13 (1):41-49.

Enthoven, A. and Singer, S., 1994, A single payer system in Jackson Hole clothing. *Health Affairs* 1: 81-95

Enthoven, A., 1991, Internal market Reform of the British National Health Service, *Health Affairs* 10: 60-70.

Enthoven, A., 1985, *Reflections of the management of the NHS.* London: Nuffield & Provincial Hospitals Trust.

European Observatory on Health Care Systems, 1999, *Health Care Systems in Transition.*

Evans, R., 1974, Supplier-Induced Demand: Some Empirical Evidence and Implications, in M. Perlman (eds.), *The Economics of Health and Medical Care*, London. McMillan.

Falkingham, J., Hemming, I., 1999, *Gender, Poverty and Access to Health Care in Tajikistan,* London School of Economics, London, U.K. Processed.

Farmer, P., 1999, Pathologies of Power: Rethinking Health and Human Rights, *American Journal of Public Health.*

Feeley, F., Sheiman, I., Shiskin, S., 1999, *Health Sector Informal Payments in Russia.* Boston University, Boston, Mass. Processed.

Feeley, F.G., V.E. Boikov, and I.M. Sheiman, 1998, *Russian Household Expenditures on Drugs and Medical Care.* Boston University, Boston, Mass. Processed.

Feldstein, M., 1976, On the theory of tax reform, *Journal of Public Economics*, v.6, pp. 77-104.

Filmer, D., Hammer, J. and Pritchett, L. 1998. Health Policy in Poor Countries. Weak Links in the Chain. *Policy Research WP 1874*, The World Bank, Washington, DC.

Filmer, D., Pritchett, L., 1997, Child Mortality and Public Spending on Health. How Much Does Money matter? *Policy Research WP 1864*, The World Bank, Washington, DC.

Filmer, D., Pritchett. L., 1998, Estimating wealth effects without expenditure data – or tears: an application to educational enrolments in the states of India, *World Bank Policy Research Working Paper*, No. 1994. World Bank, Washington.

Filmer, D., Pritchett. L., 1999, The effect of household wealth on education attainment: evidence from 35 countries, *Population and Development Review* 25: 85-120

Fox D.M., and H. Leichter, 1993, The Ups and Down of Oregon's Rationing Plan, Health Affairs, summer

1993,

Frank, G., Glazer, J., and T. McGuire, 2000, Measuring Adverse Selection in Managed Health Care, *Journal of Health Economics 19*, 6: pp.829-854.

Frenk, J., González Block, M., and Lozano, R. 1998. Seis tesis equivocadas sobre las politicas de salud en el combate a la pobreza. *Este Pais 84.*

Fuchs, V., 1978, The Supply of Surgeons and the Demand for Operations, *Journal of Human Resources 13*: 35-56.

Fudenberg, D. and J. Tirole, 1993, *Game Theory* (third edition), MIT Press, Cambridge, Mass.

Gaal, P., 1999a. *Informal Payment in the Hungarian Health Services.* Observatory Case Study (Draft). Health Services Management Training Center, Semmelweis University of Medicine, Budapest, Hungary.

Gabszewicz J., and J. Thisse, 1991, Location Models of Horizontal Differentiation: a Special Case of Vertical Differentiation Models, *Journal of Industrial Economics XXXIX*: 383-390.

Gabszewicz, J. and J. Thisse, 1996, On the Nature of Competition with Differentiated Products, *The Economic Journal: 1*60-172

Gakidou, E. Murray, C. and Frenk, G., 2000, *A Framework for Measuring Health Inequalities, Bulletin of the World Health Organization, 78*: 42-54.

Gal-or, E., Quality and Quantity Competition, The Bell Journal of Economics,

Gaynor, M., 1994, Issues in the Industrial Organization of the Market for Physician Services, *Journal of Economics & Management Strategy 3*:211-255

Gedir, G., 1999, Health Care Reform in Kyrgystan: the 'Manas Programme'. *Euro health 4*:74-76 *Georgian Economic Trends* 2001 No2 GEPLAC, TACIS, Tbilisi, 2001

Gertler, P. and J. Hammer, 1997, Strategies for Pricing Publicly Provided Health Services, in *Innovations in Health Care Financing*, G. Schieber eds., World Bank Disc. Paper n. 365.

Gertler, P. and J. van der Gaag. 1990. *The willingness to pay for medical care: Evidence from two developing countries.* Baltimore. John Hopkins Univ. Press.

Gertler, P., and J. Molyneaux. 1990. *Pricing public health services: Lessons from a social experiment in Indonesia*: RAND, Santa Monica, California.

Gilson L, Doherty J, McIntyre D, Thomas S, Brijlal V, Bowa C, Mbatsha S, 1999. The Dynamics of Policy Change: Health Care Financing in South Africa 1994-1999, Center for Health Policy School of Public Health , University of Witwatersrand

Gilson L, van den Heever A, Brijlal V, 1997. DHS Development in the North West Province : District Financing and Management , centre for Health Policy , Department of Community Health , University of Witwatersrand

Giusti, D., 2000, The fight against poverty and the private not for profit sector in Uganda. *Position paper of the Uganda Catholic, Protestant and Muslim Bureaux.*

Glazer, J., McGuire, T., 1994, Payer Competition and Cost Shifting in Health Care, *Journal of Economics and Management Strategy 3*:71-92

Glinskaya, E., J. Langenbrunner, and G. Chellaraj. 1998. *Out-of-Pocket Spending for Health Care in Russia: Who, When, Where, Why*. (Draft.) World Bank, Washington, D.C.

Golladay, F. and K. Schecter, 1998, *Ukraine: Public Expenditure Review for Health and Education*, Manuscript, World Bank.

Gonzalez-Block M, Leyva R, Zapata O, Loewe R, Alagon J, 1989. Health Services Decentralization in Mexico: Formulation, Implementation & Results of Policy. Health Policy and Planning ; 17(1) : 14-31

Gorter A, Sandiford P, Segura Z, and Villabella C. 1999. Improved health care for sex workers: A voucher programme for female sex workers in Nicaragua. In: Research for Sex Work 2. Available at: http://www.med.vu.nl/hcc/artikelen/gorter2.htm

Gotsadze G., Mataradze G. and A. Zoidze 2000. Hospital Sector Reform and Its Implications on HRD in Georgia, *Human Resources for Health Development Journal* 3 (3):185-198

Gotsadze G., Nanitashvili N. 2000, *Population's Health Needs in Western Georgia*: Guria, Imereti and Samegrelo. Curatio International Foundation. Tbilisi. p. 27

Gotsadze, G. and A. Kvitashvili., 2000, *Primary Health Care Network of Western Georgia* (Guria, Imereti and Samegrelo). Curatio International Foundation: Tbilisi.

Gotsadze, G. and Rose, L 2000, *Healthcare Financing in Georgia*, Curatio International Foundation, Tbilisi. Unpublished.

Gotsadze. G., Gzirishvili, D., Bennett. S., Ranson, K., 2001, *Health Service Utilization and Expenditures in Tbilisi – 2000, Report of a Household Survey*, Curatio International Foundation.

Government of Georgia, 2001, Georgian Economic Trends 2001, vol. 2 Tbilisi: GEPLAC, TACIS

Green A, Ali B, Naeem A, Ross D, 2000. Resource Allocation and Budgetary Mechanisms for Decentralized Health Systems: experiences from Balochistan , Pakistan. WHO Bulletin , 78(8)

Grossman, H., 1979, Adverse Selection, Dissembling and Competitive Equilibrium, *Bell Journal of Economics 10*, 336-343

Grossman,S., Hart, O., 1983, An analysis of the Principal-Agent problem, *Econometrica 51*: 7-45

Guesnerie, R., Picard, P., and P. Rey, 1989, Adverse selection and moral hazard with risk neutral agents, *European Economic Review 33*: 807-823

Gupta, S. Verhoeven, M., Tiongson, E., 2001, *Public Spending on Health Care and the Poor*, IMF Working Paper /01/127

Gwatkin, D., and Guillot, M. 1999. *The burden of disease among the global poor: current situation, future trends, and implications for strategy*. World Bank and Global Forum for Health Research, Washington.

Gwatkin, D., Guillot, M., and Heuveline, P.1999. The burden of disease among the global poor. *Lancet*, 354, 586-89.

Gwatkin D., Rutstein S, Johnson K, Pande R, Wagstaff A. 2000, *Socio-Economic Differences in Health, Nutrition, and Population in Bangladesh*. (and comparable publications covering Benin, Bolivia, Brazil, Burkina Faso, Cameroun, Central African Republic, Colombia, Comores, Côte d'Ivoire, Dominican Republic, Ghana, Guatemala, Haiti, India, Indonesia, Kenya, Kyrgyz Republic, Madagascar, Malawi, Mali, Morocco, Mozambique, Namibia, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Paraguay, Peru, Philippines, Senegal, Tanzania, Togo, Turkey, Uganda, Vietnam, Zambia, and Zimbabwe.) Washington, DC: World Bank.

Gwatkin, D., 2000a Health inequalities and the health of the poor: what do we know? what can we do? *Bulletin of the World Health Organization*, 78, 3-18.

Gwatkin, D.R. 2000b *Poverty and Inequalities in Health with Developing Countries. In: Poverty, inequality, and health* (eds. Leon, D. and G.Walt,). Oxford University Press (forthcoming)

Hammer, J. 1993. The Economics of malaria control. *The World Bank Research Observer* 8(1):1-22.

Hammer, J. 1997. Economic analysis for health projects. *The World Bank Research Observer* 12(1): 47-71

Hammer, J., Berman, P., 1995, Ends and means in public health policy in developing countries. In P. Berman, ed., *Public spending and the poor: theory and evidence.* Baltimore, Md., the J. Hopkins Univ. Press.

Hanson, K., Kikumbih, N., Armstrong Schellenberg, J., Mponda, H., Nathan, R., Mills, A., Lake, S., Tanner, M., and Lengeler, C., 2002,Cost -effectiveness of large-scale social marketing of insecticide treated nets in Tanzania. *Bulletin of the World Health Organization*, forthcoming.

Hanson K., Ranson, K., Oliviera-Cruz, Mills, A., 2001, *Constraints to scaling up health interventions: A conceptual framework and empirical analysis*, CMH WP, no WG5: 14.

Hanson, K., Berman, P., *Assessing the Private Sector: Using non-government resources to strengthen public health goals*, WP, Data for Decision Making Project, IHSG, Harvard School of Public Health.

Harris, M., Townsend, R., 1981, Resource Allocation Under Asymmetric Information, *Econometrica* 49, 33-64.

Hart, O., Holmstrom, B., 1987, The Theory of Contracts, *Advances in economic theory: fifth world congress*, T. Bewley (ed.). Cambridge: Cambridge University Press.

Hayek, F., 1978, *The Mirage of Social Justice: Law, Legislation and Liberty*, v.2 London, Routledge and Kegan Paul.

Hirschman, A., 1970, *Exit, voice and loyalty: Responses to decline in firms' organizations and states.* Cambridge Mass.: Harvard University Press.

HiT 2002, *Health Care Systems in Transition, Georgia 2001*, European Observatory on Health Care Systems, WHO Euro, Forthcoming.

Hodgkin, D. and McGuire, T., 1994, Payment levels and hospital response to prospective payment. *Journal of Health Economics*; 13: 1-29.

Hollo, I., Long, M., and A. Papp, 1998, *Health Care Financing in Hungary*, Health Financing Conference.

Holmstrom, B., 1979, Moral Hazard and Observability, *Bell Journal of Economics 10*: 74-91.

Holmstrom, B., Milgrom, P., 1990, Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design, *Journal of Law, Economics and Organization 7*: 24-52.

Hotchkiss D., Rous J., Karmacharya, K., 1998, Household Health Expenditure in Nepal; Implications for Health Care Financing Reform, *Health Policy and Planning* 13 (4): 371-383.

Independent Inquiry into Inequalities in Health (1998). HMSO, London.

ISSS 1999, *Insurance State Supervision Service of Georgia,* 1999 Annual Report.

Jacab, M., Preker, A.and A. Harding, 2001, *Hospital organization structure in transition economies: incoherent internal and external incentive environments*. (Draft). World Bank. Processed

Jack, W., 2000, Public spending on health care: how are the different criteria related? A second opinion, *Health Policy.*

Jack, W., 2000, *The evolution of health insurance institutions: theory and four examples from Latin America, Draft.*

Jacobs, P., 1974, A Survey of Economic Models of Hospitals, *Inquiry 11*, 83-97.

Jeppson A, 2001. Financial Priorities under Decentralization in Uganda Health Policy & Planning 16(2): 187-19.

Jha P, Mills A, Hanson K, Kumaranayake L, Conteh, Kurowski C, Nguyen SN, Oliveria Cruz V, Ranson K, Vaz L., Yu S, Morton O and Sachs J., 2002, Improving the health of the global poor, *Science 295*: 2036-2039.

Jost, T., Hughes, D., McHale, Jean and L. Griffiths, 1995, *The British Health Care Reforms, the American Health Care Revolution and Purchaser/Provider Contracts*, presented at the Quasi-Market Conference, London School of Economics.

Kaiser Permanente International, 1999, Georgia Hospital restructuring Master plan. Technical Report. Tbilisi

Kakwani, N., Wagstaff, A., Van Doorlsaer, E., Socioeconomic inequalities in health: measurement, computation and statistical inference, *Journal of Econometrics, 77*:87-104.

Kaplan G.., Pamuk E., Lynch J., Cohen R., Balfour J., 1996, Inequality in income and mortality in the United States: Analysis of mortality and potential pathways. *British Medical Journal* 312:999-1003.

Kawachi I, Berkman L., eds., Social cohesion, social capital, and health. In: *Social Epidemiology*, pp. 174-90. New York: Oxford University Press.

Kawachi I, Blakely T. 2001. When economists and epidemiologists disagree (commentary). *J Health Politics Policy Law* 26:533-41.

Kawachi I, Kennedy B., Lochner K, Prothrow-Stith D., 1997, Social capital, income inequality, and mortality, *American Journal of Public Health* 87: 1491-8.

Kawachi I, Kennedy B., Wilkinson R., eds. 1999. *The Society and Population Health Reader Volume 1 - Income inequality and health.* New York: The New Press, 506 pp.

Kawachi I, Kennedy B., 1999, Income inequality and health: Pathways and mechanisms, *Health Services Research* 34 (1 Pt 2): 215-227.

Kesteloot, K., Voet, N., 1998, Incentives for Cooperation in Quality Improvement Among Hospitals –the Impact of the Reimbursement System, *Journal of Health Economics 17*: 701-728.

Klugman J., Schieber, G., 1996, A Survey of Health Reform in Central Asia, *World Bank Technical Paper,* No. 344.

Klugman, J., Schieber, G., 1997, Health Reform in Russia and Central Asia. In Joan M. Nelson et al. (eds) *Transforming Post-Communist Political Economies*, National Academy Press, Washington, 1997.

Kornai, J., 2000, Hidden in an envelope. Gratitude payments to medical doctors in Hungary, *Discussion Paper Series n. 60.*

Kornai, J. and K. Eggleston, 2000, *The health sector in Eastern Europe and proposals for reform*, unpublished manuscript.

KPI, 1999, *Georgia Hospital Restructuring Masterplan. Technical Report*. Kaiser Permanente International. Tbilisi.

Labelle,R., Stoddard, G. Rice, T., 1994, A re-Examination of the Meaning and Importance of Supplier-Induced Demand, *Journal of Health Economics 13*: 347-368.

Laffont, J., Tirole, J., 1987, Auctioning incentive contracts, *Journal of Political Economy 95*: 921-937.

Laffont, J., Tirole, J., 1993, *A Theory of Incentives in Procurement and Regulation*, Cambridge MA, MIT Press

Lake S, Daura M, Mabandhla M, Masiye F, Mulenga S, Antezana I, Mwikisa C, Bennett S, 2000. *Analyzing the Process of Health Financing Reform in South Africa and in the Zambian Health Sector*, Draft, Centre for Health Policy, University of the Witwatersrand, South Africa.

Lake, S., Mtonga, S., Nakamba, P., 2002, Strengthening needs-based resource allocation

Langenbrunner, J., Wiley, M., 1999, *Paying the Hospital: payment policies and reforms. Issues, Options, Early Results*. Draft. World Bank, Washington, DC.

Lavy, V, and Strauss, J, Thomas, D., de Vreyer, P., 1996, Quality of Health Care, Survival and Health Outcomes in Ghana, *Journal of Health Economics* 15: 333-357.

Le Grand, J., 1982, *The Strategy of Equality*, London, Allen and Unwin.

Le Grand, J., 1987, Equity, Health and Health Care, *Social Justice Research*, 1:257-74.

Le Grand, J., 1991, *Equity and Choice. An Essay in Economics and Applied Philosophy*, Harper Collins, London.

Le Grand, J., Bartlett, W., 1993, *Quasi-markets and Social Policy*, London, McMillan

Leon, D. et al, 1997, Huge Variation in Russia Mortality Rates 1984-94:Artefact, Alcohol, or What?, *The Lancet 9075*: 383-388.

Leon, D., Walt,G., 2002 (eds.), *Poverty, inequality, and health*, Oxford, Oxford University Press (forthcoming).

Levine, M., Soucat, A., 2001, *HNP and the Poor. The Health System and the Poor, Session 4*. The World Bank, Washington, DC.

Lewis, M., 2002, Informal Health Payments in Eastern Europe and Central Asia: Issues, Trends and Policy Implications, in *Funding Health Care: Options in Europe*. Mossiales and Saltman (eds.) Geneva: WHO, Forthcoming.

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

Lewis, M., Eskeland, G., Traa-Valarezo, X., 1999, *Challenging El Salvador's Rural Health Strategy, DEC WP*. World Bank, Washington, DC.

Lewis, M., La Forgia, G., Sulvetta, M., 1996, Measuring Public Hospital Costs: Empirical Evidence from the Dominican Republic, *Social Science and Medicine* 43:221-234.

Lochner K, Pamuk E, Makuc D, Kennedy B., Kawachi I., 2001, State-level income inequality and individual mortality risk: a prospective, multilevel study. *American Journal of Public Health* 91:385-91.

Loevinsohn, B., 2000a, *Checklist for contracting of health services,* Draft, The World Bank, Washington, DC.

Loevinsohn, B., 2000b, *Contracting for the delivery of primary health care in Cambodia: Design and initial experience of a large pilot test, Draft.* The World Bank. Washington, DC.

Loevinsohn, B., 2001, *Contracting for Health Services: Lessons from two large projects in Bangladesh,* Draft. The World Bank. Washington DC.

Londono, B., Jaramillo, I., Uribe, J., 1999, *Decentralization and Reforms in Health Services, The Colombian Case.* Draft, The World Bank, Washington.

Lonnroth, K., 2000, *Public health in private hands. Studies on private and public tuberculosis care in Ho Chi Minh City, Vietnam.* PhD Thesis, Goteborg Department of Social Medicine and Nordic School of Public Health.

Luft, H., Robinson, J., 1985, The Impact of Hospital Market Structure on Patient Volume, Average Length of Stay and the Cost of Care, *Journal of Health Economics 4*: 333-356.

Luft, H., Robinson, J., 1987, Competition and the Cost of Hospital Care, 1972 to 1982, *Journal of the American Medical Association, 257*: 3241-3245.

Ma, C., 1994, Health Care Payment Systems: Cost and Quality Incentives, *Journal of Economics and Management Strategy 3*:93-112.

Ma, C., McGuire, T., 1997, Optimal Health Insurance and Provider Payment, *American Economic Review (Sept.)*: 685-700.

Maceira, D., 1998, Income Distribution and the Public-Private Mix in Health Care Provision: the Latin American Case, Inter American Development Bank WP.

Mahal, A., Yazbeck, A., Peters, D., 2001, *The Poor and Health Service Use in India.* Draft. The World Bank, Washington, DC.

Maly, I. and P. Veprek, 1999, The Managed Care with a Human Face? (An Implementation of Managed Care Principles to the Czech Health Care System, *(Draft).*

Manning W, Newhouse, J., Duan, N., Keeler, B., Leibowitz, A., and M. Marquis, 1987, Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *American Economic Review* 77: 251-77.

Manning, N., 1998. *Unbundling the State: Autonomous Agencies and Service Delivery.* Draft. Washington, D.C.: Economic Development Institute, World Bank, April 29.

Marek T, Diallo I, Ndiaye B, Rakotosalama J. 1999. Successful contracting of prevention services: fighting malnutrition in Senegal and Madagascar. *Health Policy and Planning* 1999 14: 382-389.

Marquis, M., 1992, Adverse Selection with a Multiple Choice Among Health Insurance Plans: A Simulation Analysis, *Journal of Health Economics 11*, 129-151.

Martimort, D., 1991, Multi-principaux Avec Selection Adverse. WP, *Istitut d'Economie Industrielle,* Toulouse.

Matsaganis, M., Glennester, H., 1994, The Threat of Cream Skimming in the Post-Reform NHS, *Journal of*

Health Economics 13: 31-60.

Maynard, A., 1991, Developing the Health Care Market, *Economic Journal 101*, 1277-86.

Mays J., Schaefer M., 1999. *Health Financing Study of Georgia: Impacts of Alternative Prototypical Options*, Actuarial Research Corporation.

McAfee R., McMillan, J., 1986, Bidding for contracts: a principal agent analysis, *RAND Journal of Economics 17*: 326-338.

McAfee, R., McMillan, J., 1987, Competition for Agency Contracts, *RAND Journal of Economics 18/2*.

McAfee, R., McMillan, J., 1988, Multidimensional Incentive Compatibility and Mechanism Design, *Journal of Economic Theory 46*, 335-354.

McGuire, T., Riordan, M., 1994, Introduction: The Industrial Organization of Health Care, *Journal of Economics and Management Strategy 3*, 1-6.

McGuire, T., Riordan, M., 1995, Incomplete Information and Optimal Market Structure. Public Purchases form Private Providers, *Journal of Public Economics 56*, 125-141.

McIntyre, D., Gilson, L., 2002, Putting equity in health back onto the social policy agenda: experience from South Africa. *Social Science and Medicine, 54*; 1637-1656, 2002.

McKee, and Leon, Understanding the Russian Mortality Crisis, *Eurohealth. 3*.

McPake B, Yepes FJ, Lake S and Sanchez LH. 2002. Is the public health system reform improving the performance of public hospitals in Bogotá? Manuscript.

Metrick, A. and R. Zeckhauser, 1996, *Price versus Quantity: Market Clearing Mechanisms When Sellers Differ in Quality*, NBER WP 5728.

Mills A, Bennett S, Siriwanarangsun P and Tangcharoensathien V., 2000, The response of providers to capitation payment: a case-study from Thailand. *Health Policy 51*: 163-180.

Mills A, Brugha R, Hanson K and McPake B., 2002, What can be done about the private health sector in low-income countries? *Bulletin of the World Health Organization, 80(4)*: 325-330.

Mills, A., 1998, To contract or not to contract? Issues for low and middle-income countries, *Health Policy and Planning 13*: 32-40.

Mills, A., Bennett, S. and McPake, B. (eds.) *Private Health Providers in Developing Countries: Serving the Public Interest?* Zed Books, 1997.

Ministry of Health, Labor and Social Affairs 2000, *Annual Report 1999 of The Ministry of Health of Georgia*, 2000, Tbilisi, p. 9.

MoLHSA 1998–1997, *Annual Report of The Ministry of Health of Georgia*, 1998, Tbilisi, p. 26.

Miranda, E., Paredes, R., 1998, Competition, Vertical Integration and Performance in Chile's Public and Private Health Services", Chapter 6 in W. Savedoff, ed. *Organization Matters: Agency Problems in Health and Education in Latin America*, Latin American Research Network, Inter-American Development Bank, Washington, DC.

Mitchell, J., Scott, E., 1992, Physician Ownership of Physical Therapy Services: Effects on Charges, Utilization, Profits and Service Characteristics, *Journal of the American Medical Association 268*: 2055-2059

Miyazaki, H., 1977, The Rat Race and Internal Labor Markets, *Bell Journal of Economics 8*, 394-418.

Monigsbaun, F., 1992, *Who Shall Live, Who Shall Die? Oregon's Health Financing Proposal*. King's Fund College.

Mooney, G., Ryan, M., 1993, Agency in Health Care: Getting Beyond First Principles, *Journal of Health Economics 12*: 125-136.

Mooney, G., 1986, *Economics, Medicine and Health Care*. Brighton: Wheatsheaf.

Mossialos E, Dixon A, Figueras J, Kutzin J, editors, 2002, Funding Health Care. European Observatory on Health Care Systems Series. Open University Press. Geneva, WHO.

Motta, M., 1993, Endogenous Quality Choice: Price vs. Quantity Competition, *Journal of Industrial Economics XLI*: 113-131.

Murray C., Lopez, A., 1994, Global and Regional Cause-of-Death Patterns in 1990. *Bulletin of the WHO, Geneva 72/3*: 447-480.

Murray C, Lopez A., (eds.), 1996, *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries and Risk Factors in 1990 and Projected to 2020*. Cambridge, Harvard School of Public Health on behalf of the WHO and the World Bank (Global Burden of Disease and Injuries Series, Vol. 1).

Murray, C. and A. Lopez, 1997, Global Mortality, Disability, and the Contribution of Risk Factors: Global Burden of Disease Study," *The Lancet*, v. 349, May 17, pp. 1436-42.

Murray, C., Gakidou, E., and Frenk, J., 1999, Health inequalities and social group differences. *Bulletin of the World Health Organization 77*, 537-44.

Musgrave, R. , 1976, ET, OT and SBT. *Journal of Public Economic 6(1):* pp.3-16.

Musgrave, R., 1996, The Role of the State in Fiscal Theory, *International Tax and Public Finance 3*: 247-58

Musgrove, P. 1999. Public Spending on health care: how are different criteria related? *Health Policy 47*: 207-223.

Musgrove, P., 1996, Public and Private Roles in Health: Theory and Financing Patterns, http://www1.worldbank.org/hnp/HNP%20Pubs%20-%20Discussion/Musgrove%20public-private-whole.pdf

Musgrove, P., 1996, Un fundamento conceptual para el rol público y privado en salud" *Revista de Analisis Economico*, Vol. 11:2, November.

Mwabo, G., Ainsworth, M., Nyamete, A., 1993, Quality of Care and Choice of Medical Treatment in Kenya. An Empirical Analysis, *Journal of Human Resources 28*: 838862.

Nakamba, Hanson and McPake. 2002. Markets for hospital services in Zambia. *International Journal of Health Planning and Management 17*: 229-247.

Narayan, D. et al., 1999, *Global Synthesis: Consultation with the Poor*, World Bank: Washington DC.

Neudeck, W., Podczeck, K, 1996, Adverse selection and Regulation in Health Insurance Markets, *Journal of Health Economics 15*: 387-407.

Newhouse, J., 1984, Cream Skimming, Asymmetric Information and a Competitive Insurance Market, *Journal of Health Economics 3*, 97-100.

Newhouse, J., 1989, Do unprofitable patients face access problems? *Health Care Financing Review 11*: 2. Pp. 33-42 EMBASE.

Newhouse, J., 1993, *Free for all?: Lessons from the Rand health insurance experiment.* Cambridge Mass.: Harvard University Press.

Newhouse, J., 1996, Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection, *Journal of Economic Literature 34*, 1236-1263.

Noether, M., 1988, Competition Among Hospitals, *Journal of Health Economics 7*: 259-284.

Nozick, R., 1974, *Anarchy, State and Utopia.* Basic Books. New York.

Okun, A. 1975, *Equality and Efficiency: the Big Trade-off.* Washington, DC: Brookings.

Olsen, E and Rogers, D., 1991, The welfare economics of equal access, *Journal of Public Economics 45*: 91-106.

Orosz, E. and Hollo, 1999, I. *Hospital Sector in Hungary -The story of unsuccessful reforms.* (Draft) World Bank, Washington, D.C.

Palmer, N., 2000, The use of private-sector contracts for primary health care: theory, evidence and lessons for low-income and middle-income countries. *Bulletin of the World Health Organization 78(6)*: 821-30.

Palmer, N., 2001, *Alternative contract designs for private providers of primary health care: a case study of South Africa.* PhD thesis, University of London.

Palu, T., 1999, *Hospital Sector in Transition Baltics Case Study.* (Draft.) World Bank, Washington, D.C.

Pande, R., Yazbeck, A., 2001, *Beyond National Averages for Immunization: Income, Gender, and Regional Inequalities*, HNP Working paper, The World Bank, Washington, DC.

Panis, C., Lillard, L., 1994, Health inputs and child mortality: Malaysia, *Journal of Health Economics* 13: 455-489.

Pauly, M., 1994, Editorial: A Re-examination of the Meaning and Importance of Supplier-Induced-Demand, *Journal of Health Economics 13*: 369-372.

Peters, D., Yazbeck, A., Sharma, R., Ramana, GNV, Pritchett, L. and A. Wagstaff, 2002, Better Health Systems for India's Poor, The World Bank, Washington D.C.

PIEDR (Pakistan Institute for Environment-Development Action) and Project Management Team, 1994, *The State of the Public Sector Primary Health Services, District Sheikhupura, Punjab, Pakistan,* Bamako initiative technical report series; n. 31, UNICEF.

Preker, A., A. Harding, and N. Girishankar, 1999, *The Economics of Public and Private Participation in Health Care: New Insights from Institutional Economics.* (Draft.) Paper presented at The Economist Forum, Alexandria, VA, May 4, 1999. World Bank, Washington, D.C.

Preker, A., Dror, D., eds., 2002, Social Reinsurance: A New Approach to Sustainable Community Health Financing, ILO, World Bank.

Preker A, Jakab M, Shneider M., Health financing reforms in central and eastern Europe and the former Soviet

Newhouse, J., 1984, Cream Skimming, Asymmetric Information and a Competitive Insurance Market, *Journal of Health Economics 3*, 97-100.

Newhouse, J., 1989, Do unprofitable patients face access problems? *Health Care Financing Review 11*: 2. Pp. 33-42 EMBASE.

Newhouse, J., 1993, *Free for all?: Lessons from the Rand health insurance experiment.* Cambridge Mass.: Harvard University Press.

Newhouse, J., 1996, Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection, *Journal of Economic Literature 34*, 1236-1263.

Noether, M., 1988, Competition Among Hospitals, *Journal of Health Economics 7*: 259-284.

Nozick, R., 1974, *Anarchy, State and Utopia.* Basic Books. New York.

Okun, A. 1975, *Equality and Efficiency: the Big Trade-off.* Washington, DC: Brookings.

Olsen, E and Rogers, D., 1991, The welfare economics of equal access, *Journal of Public Economics 45*: 91-106.

Orosz, E. and Hollo, 1999, I. *Hospital Sector in Hungary -The story of unsuccessful reforms.* (Draft) World Bank, Washington, D.C.

Palmer, N., 2000, The use of private-sector contracts for primary health care: theory, evidence and lessons for low-income and middle-income countries. *Bulletin of the World Health Organization 78(6)*: 821-30.

Palmer, N., 2001, *Alternative contract designs for private providers of primary health care: a case study of South Africa.* PhD thesis, University of London.

Palu, T., 1999, *Hospital Sector in Transition Baltics Case Study.* (Draft.) World Bank, Washington, D.C.

Pande, R., Yazbeck, A., 2001, *Beyond National Averages for Immunization: Income, Gender, and Regional Inequalities*, HNP Working paper, The World Bank, Washington, DC.

Panis, C., Lillard, L., 1994, Health inputs and child mortality: Malaysia, *Journal of Health Economics* 13: 455-489.

Pauly, M., 1994, Editorial: A Re-examination of the Meaning and Importance of Supplier-Induced-Demand, *Journal of Health Economics 13*: 369-372.

Peters, D., Yazbeck, A., Sharma, R., Ramana, GNV, Pritchett, L. and A. Wagstaff, 2002, Better Health Systems for India's Poor, The World Bank, Washington D.C.

PIEDR (Pakistan Institute for Environment-Development Action) and Project Management Team, 1994, *The State of the Public Sector Primary Health Services, District Sheikhupura, Punjab, Pakistan,* Bamako initiative technical report series; n. 31, UNICEF.

Preker, A., A. Harding, and N. Girishankar, 1999, *The Economics of Public and Private Participation in Health Care: New Insights from Institutional Economics.* (Draft.) Paper presented at The Economist Forum, Alexandria, VA, May 4, 1999. World Bank, Washington, D.C.

Preker, A., Dror, D., eds., 2002, Social Reinsurance: A New Approach to Sustainable Community Health Financing, ILO, World Bank.

Preker A, Jakab M, Shneider M., Health financing reforms in central and eastern Europe and the former Soviet

Union. In: Mossialos E, Dixon A, Figueras J, Kutzin J, editors. Funding health care. European Observatory on Healthcare Systems Series. Open University Press; 2002. p. 80–109.

Preston, S. 1980. *Causes and consequences of mortality declines in less developed countries in the 20th century*. In: Easterlin, R. (ed.), *Population and Economic Change in Developing Countries*. Chicago, Univ. of Chicago Press: 289-341

Pritchett, L , Summers, L., 1996. Wealthier is healthier. *Journal of Human Resources* 31(4): 841-868.

Propper, C., 1996, Market structure and prices: The responses of hospitals in the U.K. National Health Service to Competition, *Journal of Public Economics 61*, 307-335

Propper, C., Wilson, D. and N. Soderlund, 1997, *The Effects of Regulation and Competition in the NHS Internal Market: The Case of GP Fund-Holder Prices*, presented at the European Economic Ass. Conference in Toulouse, France.

Rajani, R., 2001, *The Education Pipeline in East Africa*, Draft.

Ravallion, M., Wodon, Q., 1997, Poor areas, or only poor people? *Policy Research Working Paper 1798*, Development Research Group, The World Bank, Washington DC.

Rawls, J. 1971, *A Theory of Justice*, Cambridge, MA., Belknap Press.

Reagon, G, Makan, B, McIntyre D., 1997, *Resource Allocation Mechanisms within the District Health System: A Literature Review*, Health Economics Unit, department of Community Health , University of Cape Town

Resource Allocation Working Party, London, HMSO, DHSS.

Rhodes G 2000, *Rapid Assessment of Co-payments Options in Georgia*, Unpublished paper prepared for the Health Fund Development Project.

Rice, N. and Smith, P., 1999, *Own Experience with Resource Allocation Formulae in North, Approaches to Capitation and Risk Adjustment in Health Care: an International Survey*, Centre for Health Economics, Occasional Paper.

Rice, N., and Smith, P., 2000, *Strategic Resource Allocation and Funding Decisions*, Paper presented at the European Observatory on Health Care Systems' Project on Funding Health Care: Options for Europe.

Riley, J., 1979, Informational Equilibrium, *Econometrica 47*, 331-360.

Rodwin V., 2000, *On the Separation of Health Planning and Provider Reimbursement*: The U.S. and France. http://www.nyu.edu/projects/rodwin/separation.html

Roemer, M., 1961, Bed Supply and Hospital Utilization: A natural Experiment, *Hospitals 35*: 36-42

Rogerson, W., 1994, Choice of treatment intensities by a non profit hospital under prospective pricing, *Journal of Economics and Management Strategy 3*:7-51.

Ronnen, U., 1991, Minimum Quality standards, fixed costs, and ompetition, Rand Journal of Economics 22-4: 490503.

Rothschild, M., Stiglitz, J., 1976, Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, *Quarterly Journal of Economics 90*: 630-649.

Rothschild, M., Stiglitz, J., 1997, Competition and Insurance Twenty Years Later, *The Geneva Papers on Risk and Insurance Theory 22*, 73-79.

Russell, L., 1989, *Medicare's New Hospital Payment System: Is It Working?* Washington: The Brookings Institution.

Ruzica, M. et al., 1999, *Moldova Health Reform: Social and Institutional Assessment (SIA)*. (Draft.) World Bank, Washington, D.C.

Saltman. R., Figueras. J., and Sakellarides. C., 1998, *Critical Challenges for Health Care Reform in Europe.* Open University Press.

Sánchez F., Núñez, J., 1999, *La Ley 100/93 Sí Llegó a las Personas de Menores Ingresos"*, *La Revista VíaSalud*, third trimester, Santafé de Bogotá, Colombia.

Sánchez, H. and G. Zuleta, eds. 2000. *La hora de los usuários: Reflexiones sobre economía política de las reformas de salud*. InterAmerican Development Bank, Washington, DC.

Sapelli, C. and B. Vial, 1998, Utilización de prestaciones de salud en Chile: Es diferente entre grupos de ingreso? *Cuadernos de Economía* 35, No. 106, pp. 343-382.

Sargent, T. and N. Wallace, 1985, *Some Unpleasant Monetarist Arithmetic*, Quarterly Review Vol. 5 No. 3.

Sari, A., Langenbrunner, J., and M. Lewis, 2000, Affording Out-of-Pocket Payments for Health Care Services: Evidence from Kazakhstan, *Eurohealth* 6,2 (Spring).

Savedoff, W., 1998, Social Services Viewed Through New Lenses, Chapter 1 in: Savedoff, W. (ed): *Organization Matters: Agency Problems in Health and Education in Latin America*, Latin American Research Network, Inter-American Development Bank, Washington, DC.

Savedoff, W., 2001, *Is Anybody Listening? Ignoring Evidence in the Latin American Health Reform Debates. A Health Note. SDS/SOC, Health Note n 2.* Inter-American Development Bank, Washington, DC

Schaapveld, K., 2000, *The Georgian State Medical Insurance Company (SMIC) in 1999*. Wassenaarseweg, Leiden. TNO Prevention & Health.

SDS, 2000ᵃ, *Georgia's Households 1998-1999.* Georgia State Department of Statistics, Tbilisi.

SDS, 2000ᵇ, *Medical Service in Georgia, Survey Report on Population Health-care Expenses and Unrecorded Medical Service*. Georgia State Department of Statistics, Tbilisi.

Secretary of State of England, 1997, *The New NHS*. London, HMSO.

Sen, A., 1992, *Inequality reexamined*, Oxford: Oxford University Press.

Sen, A., 1997a, Editorial: *Human capital and human capability*, World Development, 25, 12: 1959-1961.

Sen, A., 1997b, Health and Poverty in the context of development strategy: a case study of Bangladesh. *Macroeconomics, Health, and Development Series*, No. 26. WHO, Geneva.

Sen, A., 1973, *On the development of basic income indicators to supplement GNP measures*, UN Economic Bulletin for Asia and the Far East, 24, 2-3: 1-11.

Sen, A., 1984, *Resources, Values and Development*. Cambridge MA, Harvard University Press.
Sen, A., 1985, *Commodities and capabilities*, Amsterdam: North Holland.

Sen, A., 1989, *Food and Freedom*, World Development, 17: 769-81.

Sen, A., 1990, Individual freedom as a social commitment, *The New York Review of Books*, 37, 10: 49-54

Sen, A., 1993, Capability and Well-being, in Nussbaum, M., and Sen, A. (eds.) *The Quality of Life* (Oxford: Clarendon Press for WIDER/UN): pp. 30-53.

Sen. A., 1994, The Political Economy of Targeting. In: *Public Spending and the Poor: Theory and Evidence* (ed. van de Walle, D., Nead, K.), pp. 11-24. Johns Hopkins University Press.

Shahriari, H., Belli, P. and Lewis, M., 2001, *Informal Payment in Health in Poland.* The World Bank, Washington D.C..

Shaked, A., Sutton, J., 1983, Natural Oligopolies, *Econometrica 51(5)*: 1469-1483.

Shaked, A., Sutton, J., 1987, Product Differentiation and Industrial Structure, *Journal of Industrial Economics XXXVI*: 131-145.

Shkolnikov, V., 1997, The Russian Health Crisis of the 1990s in Mortality Dimensions, *WP 97, Harvard Center for Population & Development Studies.*

Shleifer, A., 1985, A Theory of Yardstick Competition, *RAND Journal of Economics 16*: 319-327.

Slack K., Savedoff, W., 2000, *Public purchaser-private provider contracting for health services: examples from Latin America and the Caribbean.* Washington (DC): Inter-American Development Bank (SDS Technical Paper).

Smith E, Brugha R and Zwi A. 2001. *Working with private sector providers for better health care, an introductory guide.* London: Options & London School of Hygiene and Tropical Medicine.

Soucat, A, Yazbeck, A., 2001, *NHP and the Poor: Inputs into PRSPs and Bank Operations (2001): Session 1.* World Bank, Washington DC.

Soucat, A., Wagstaff, A., 2001, *HNP and the Poor: An Integrated Policy Framework for Improving the Outcomes for the Poor, Session 6.* World Bank, Washington DC.

Spence, M., 1978, Product Differentiation and Performance in Insurance Markets, *Journal of Public Economics 10*, 427-447.

Ssengooba F, Atuyambe L, McPake B, Hanson K and Okuonzi S. (Forthcoming.) What could be achieved with greater public hospital autonomy? Comparison of Public and PNFP Hospitals in Uganda. *Public Administration and Development.*

State Department of Statistics, (SDS), Medical Services In Georgia: survey report on population health care expenses and unrecorded medical services. Tiblisi, Georgia State Department of Statistics

Stiglitz, J., 1977, Monopoly, Non-linear Pricing and Imperfect Information: The Insurance Market, *Review of Economic Studies 44*, 407-428.

Stiglitz, J., 1999, *Incentives and Institutions in the Provision of Health Care in Developing Countries: towards an Efficient and Equitable Health Care Strategy*, presented at the iHEA II Meetings, Rotterdam, June.

Stout, S., Johnston, T., 1999, *Investing in Health Development Effectiveness in the Health, Nutrition, and Population Sector.* World Bank, Washington, DC.

Streeten, P., 1994. Human Development: Means and Ends, *The American Economic Review* 84, 2: 232-237.

Streeten, P., Burki, S., 1978, Basic needs: some issues, *World Development* 6 (3): 411-421.

Streeten, P., Burki, S., Haq, M., Hicks, N., and Stewart, F, 1981, *First Things First, Meeting Basic Human*

Needs in the Developing Countries. New York: Oxford University Press for The World Bank.

Subramanian S., Kawachi I, Kennedy B., 2001. Does the state you live in make a difference? Multilevel analysis of self-rated health in the U.S.. *Social Science and Medicine* 53.

Subramanian, S., Belli, P., Kawachi, I., 2001, The Macroeconomic Determinants of Health, Annual Reviews, *American Review of Public Health.*

Swedish Parliamentary Priorities Commission, 1995, *Priorities in Health Care: Ethics, economy, implementation*, Fritzes, Stockholm.

Taylor, R., 2000, Contracting for Health Services, Background paper, *Private Participation in Health Handbook*, The World Bank, Washington DC

Thatcher, M., 1989, Foreword to *Working for Patients: The Health Service, Caring for the 1990s*. HMSO, London.

Tirole, J., 1988, *The Theory of Industrial Organization*, Cambridge, MA: MIT Press

Tobin, J., 1970, On Limiting the Domain of Inequality, *Journal of Law and Economics* 13: pp. 263-77.

Trombetta, M., 1997, *Externalities and International Minimum Quality Standards when Demand is Rigid*, Ph.D. University of Oxford.

UNDP, 1997, *Poverty and Human Development* New York: Human Development Report Office, The United Nations Development Program..

UNICEF, 1993, *Georgia women's health profile.*

Valletti, T., 1998, *Minimum Quality Standards under Cournot Competition*, PhD., The London School of Economics.

Van de Ven, W., Ellis, R., 2000, Risk-adjustment in competitive health plan markets, in Newhouse, J., Culyer, A. (eds..) *Handbook of Health Economics*, North Holland, Elsevier, Amsterdam

Van de Ven, W., et al. 2000, Access to coverage for high-risks in a competitive individual health insurance market: via premium rate restrictions or risk-adjusted subsidies? *Journal of Health Economics 19,3*: pp. 311-339.

Van de Ven, W., Van Vliet, R., 1992, *How Can we Prevent Cream Skimming in a Competitive Health Insurance Market? The Great Challenge for the 90's*, in Zweifel, P., Frech, H. (III eds), *Health Economics Worldwide*, Kluwer Amsterdam, 23-26.

Van de Ven, W., Van Vliet, R., 1995, Consumer Information Surplus and Adverse Selection in Competitive Health Insurance Markets: an Empirical Study, *Journal of Health Economics 14*, 149-169.

Van de Walle, D., Nead, K. 1998. *Public Spending and the Poor*, The World Bank, Washington.

Van der Gaag, J. 1995. *Private and public initiatives: working together for health and education. Directions in development Series*. The World Bank, Washington, D.C.

Van Doorslaer, E. et al., 1997, Income-related inequalities in health: Some international comparisons. *Journal of Health Economics* 16: 93-112.

Van Doorslaer, E. et al. 2000, Equity in the delivery of health care: further international comparisons, *Journal of Health Economics*.

Van Doorslaer, E., Wagstaff, A et al. 1997. Equity in Health Care Financing: Redistributive Effects and Horizontal Inequality. *Equity Project Working Paper # 8*. Rotterdam. Erasmus University.

Van Doorslaer, E., Wagstaff, A, Rutten, F., 1993, *Equity in the finance and delivery of health care: An international perspective.* Commission of the European Communities, Health Service Research Series, N. 8, Oxford. Medical Publications.

Van Zanten, Veldhuyzen.T., and Semidei, C., 1996. *Assessment of Health Sector Decentralization in Paraguay.* Technical Report No. 1. Bethesda, MD: Partnerships for Health Reform Project, Abt Associates Inc.

Varian, H, 1984, *Microeconomic Analysis*, 2nd. Ed. New York, W.W. Norton.

Vickers, J., Yarrow, G., 1988, *Privatization: an Economic Analysis*, Cambridge, MA, MIT Press

Victora, C. et al., 2000, Explaining trends in inequities: evidence from Brazilian child health studies. *Lancet* 356: 1093-98

Wagstaff A , van Doorslaer, E., 1998. *Inequalities in health: methods and results for Jamaica.* Paper prepared for the Human Development Department of the World Bank. Second version May 1998.

Wagstaff, A. 1999, *Inequalities in Infant and Child Mortality in the Developing World: How Large Are They? How Can They Be Reduced?* Presentation to World Bank HNP and Poverty Seminar, March 30.

Wagstaff, A., 2000a, *Research on Equity, Poverty and Health: Lessons for the Developing World*, Draft. World Bank, Washington D.C.

Wagstaff, A., 2000b, Socioeconomic inequalities in child mortality: Comparisons across nine developing countries. *Bulletin of the World Health Organization*, 78, 19-29.

Wagstaff, A., 2001, *Poverty and Health, paper presented for the Commission on Macroeconomics and Health, WHO,Geneva*

Wagstaff, A., Paci, P., Van Doorslaer, E. 1991. On the measurement of horizontal equity in the delivery of health care. *Journal of Health Economics* 10: 169-205.

Wagstaff, A., Paci, P., Van Doorslaer, E. 1991. On the measurement of inequalities in health, *Social Science and Medicine* 33: 1545-557.

Wagstraff, A.,Van Doorslaer,E., 2000, *Equity in Health Finance and Delivery*, Chapter 40, *Handbook of Health Economics*, Elsevier Science.

Wagstaff, A., and Yazbeck, A., 2001, *The Roles and Constraints of Households and Communities, The Health System and the Poor, Session 4*. The World Bank, Washington, DC.

Wallace R, Wallace D., 1997, Community marginalisation and the diffusion of disease and disorder in the United States. *British Medical Journal* 314:1341-5.

Waters, H., 1995, *Literature review: equity in the health sector in developing countries, with lessons learned for Sub-Saharan Africa.* Unpublished manuscript.

Waters, H., 2000, Measuring Equity in Access to Health Care, *Social Science and Medicine* 51: 599-612.
Wauthy, X. 1996, Quality Choice in Models of Vertical Differentiation, *Journal of Industrial Economics XLIV*: 345-359.

Way, L. 1999, *Decentralization of Social Services in Ukraine.* (Draft.) San Francisco.

Way, L., Belli, P., Lewis, M. et al., 1999, *Proposal for Research on informal payments in health in ECA.*, (Draft.) World Bank, Washington, D.C..

Weil, D., 2000, Advancing Tuberculosis Control within Reforming Systems, *The International Journal of Tuberculosis and Lung Disease 4: 597-605*

Wilkinson R., Kawachi I, Kennedy B., 1998, Mortality, the social environment, crime and violence. *Sociology of Health and Illness* 20: 578-97.

Wilkinson R., 1996. *Unhealthy Societies: The Afflictions of Inequality.* London: Routledge.

Wilkinson R., 2000. *Mind the Gap: Hierarchies, Health and Human Evolution.* London: Weidenfeld & Nicolson.

Wilkinson, R., 1997, Health inequalities: relative or absolute material standards? *British Medical Journal* 314: 591-595.

Wilson, C., 1977, A Model of Insurance Market with Incomplete Information, *Journal of Economic Theory 16:* 167-207.

World Bank, Washington, D.C.:

———1993. *World Development Report.*

———1996a *Georgia Public Expenditure Review*, Report No. 15779.

———1996b, *Staff Appraisal Report No. 15069-*Georgia Health Project.

———1997a. Sector Strategy: Health, Nutrition and Population.

———1997b. *Russian Federation, Health Sector Strategy Note (Draft).*

———1998a *Health Strategy Paper for Eastern Europe and Central Asia.*

———1998b. *Reducing Poverty in India: Options for more effective public service.*

———1999a. *Czech Republic Country Economic Memorandum.*

———1999b. *Georgia: Poverty and Income Distribution.* Report No. 19348-GE, Vols. I/II.

———1999c. *Under-the-Table Payment and Health Care Reforms in Hungary.* Observatory Case Study (Draft). Health Services Management Training Centre, Semmelweis University of Medicine, Budapest, Hungary.

———1999d. *Russia: TB/AIDS Project. Aide Memoire* I.

———2000a. *"Health"* Chapter in *Czech Republic: Public Expenditure Review.* (Draft.)

———2000b. *Making Transition Work for Everyone: Poverty and Inequality in Europe and Central Asia*

———2000c. *World Development Indicators.*

———2000d. *ECA Live Database*

———2001a. *HNP Chapter for PRSP Sourcebook.*

————2001b. *Georgia Poverty Update*, Report No. 22350

————2001c *World Development Indicators.*

————2002a, *Georgia Public Expenditure Review, The Health Sector (Chapter 5).*

————2002 *World Development Indicators.*

World Health Organization, Geneva:

————1978. *Primary health hare: report of the international conference on primary health care*, Alma-Ata, USSR, 6–12 September 1978.

————1995, *World Health Report 1995: Bridging the Gap.*

————1996a. *Equity in Health and Health Care: a WHO/SIDA Initiative.* Document WHO/ARB/961.

————1996b. *Health Care Systems in Transition*, Estonia.

————1998. *Health Care Systems in Transition*, Poland.

————1999a, *World Health Report: Making a Difference.*

————1999b, *Poverty and Health*: Report by the Director General, document EB105/5 presented to the 105th session of the WHO Executive Board.

————2000. *World Health Report: Health Systems: Improving Performance.*

————2001a, *Health for All Database.*

————2001b. *Involving private practitioners in tuberculosis control: issues, interventions and emerging policy framework* (unpublished document WHO/CDS/TB/2001.285; available from: URL: http://whqlibdoc.who.int/hq/2001/WHO CDS TB 2001.285.pdf

Yazbeck, A., Sherma, R., Ramana, G., 2001, *Determinants of Immunization Coverage by Socio-economic Group in India*, draft., World Bank, Washington, D.C.

Zatonski, W., Hulanicka B., Tyczynski J., eds., 1996, *Stan Zdrovia Polakow* (The health status of the Poles). Monografie Zakladu Antropologii Polskiej Akademii Nauk, Wroklaw.

Zoidze, A., Gzirishvili, D., Gotsadze, G., 1999, Hospital Financing Study for Georgia, SAR4 Bethesda, MD: Partership for Health Reform Project, Abt. Associates Inc.