Theory and evidence on cost sharing in health care: *an economic perspective*

Theory and evidence on cost sharing in health care: *an economic perspective*

Martin Chalkley
and
Ray Robinson

Office of Health Economics
12 Whitehall London SW1A 2DY

ISBN 1899040 25 0

Printed by BSC Print Ltd., London

About the authors

Dr Martin Chalkley is Lecturer in Economics at the University of Southampton.

Ray Robinson is Professor of Health Policy and Director of the Institute for Health Policy Studies at the University of Southampton.

Office of Health Economics

The Office of Health Economics was founded in 1962 by the Association of the British Pharmaceutical Industry. Its terms of reference are:

To undertake research on the economic aspects of medical care.

To investigate other health and social problems.

To collect data from other countries.

To publish results, data and conclusions relevant to the above.

The Office of Health Economics welcomes financial support and discussions on research problems with any persons or bodies interested in its work.

For recent OHE publications see inside back cover.

Contents

1 Introduction

Faced with the growing pressures on health care budgets, policy makers around the world have turned to different forms of direct charging for health services. However, because it is rare to find a health system where the user is faced with the full cost of the service, these charges are often referred to as *cost sharing*. That is, the cost of the service is shared between the user and some third party payer, typically a sickness fund, insurance company or government agency.

Various reasons have been cited for the use of cost sharing arrangements, some of them contradictory (Evans, Barer and Stoddart, 1995). One motivation has been to raise revenues to offset escalating health care expenditures, in both the public and private sectors. Others have argued that charges represent a useful supplementary source of revenue – in addition to public finance – for funding health care. However, from an economist's point of view, the main focus of interest is on the impact that charges have upon health service utilisation. In particular, do user charges make consumers more aware of the cost of the services that they receive and thereby discourage excessive or unnecessary use?

In the UK cost sharing has long been a feature of the pharmaceutical, dental and ophthalmic services. From time to time, its more general application is sometimes canvassed in relation to, for example, payments for GP consultations and/or charges for the 'hotel' element of hospital inpatient stays. At the time of writing, the scope for using charges in this way has been widely reported as part of the new government's comprehensive review of NHS spending. The purpose of this paper is: (i) to provide a framework in which the arguments for and against cost sharing can be understood (ii) to provide an overview of the extent of cost sharing in practice and (iii) to consider the evidence regarding its effects. Our perspective on these issues is essentially an economic one.

Accordingly, the paper is organised as follows:

- Section 2 considers the economics of cost sharing. This involves a brief discussion of the key characteristics of health care that have led to its distinctive form of financing and the presentation of an evaluative framework that can be used to assess the performance of cost sharing and other policy instruments;

- Section 3 discusses the different forms that cost sharing may take, with particular emphasis on the economic incentives offered by different payment systems;

- Section 4 provides some summary data indicating the form and extent of cost sharing in major OECD countries and a brief review of the way that it is used in five of these countries;

- Section 5 reviews the evidence regarding the effects of cost sharing. This review considers the international evidence but places particular emphasis on one well documented experiment; namely, the RAND Health Insurance Experiment (Manning *et al.*, 1988). While limited in some respects, this study undoubtedly provides the most reliable available evidence on the effects of cost sharing in practice. In this section we also review the empirical evidence relating to cost sharing and pharmaceuticals in the UK;

- The final section of the paper summarises the discussion and reflects on practical policy conclusions.

2 The economics of cost sharing

To appreciate the issues that arise in the context of cost sharing for health services, it is first necessary to understand some of the basic principles that guide the economic analysis of health care in general. Accordingly, in this section we outline two key features of health services that distinguish them from other goods and services and consider the implications of these distinguishing features for health care policy both as it relates to cost sharing and other policy measures.

2.1 Two characteristics of health services and markets

The first characteristic to emphasise in relation to health care is *imperfect information*. The fact that consumers of health services typically have a very poor perception of the potential benefit of particular services compared with doctors has been viewed as an important part of the economic analysis of health care since Arrow's (1963) seminal paper. Since then there has been a growing interest in economics generally in the implications of imperfect information for the functioning of resource allocation mechanisms, whether private (market) based or public (see, for example, Mass-Collel *et al.*, 1995). This literature has shown that imperfect information affects much more than patients' ability to judge the quality, effectiveness and value of the health services that they might be offered. Whenever health services are purchased on behalf of patients by a third party purchaser that purchaser also labours under imperfect information regarding exactly what is being delivered by way of services and their effectiveness. There are also, or course, information imperfections on the supply-side resulting from some doctors' lack of knowledge of best practice and genuine uncertainties surrounding the effectiveness of some clinical practice. These considerations bring us to the second characteristic of health services.

In most health care systems, patients are not faced directly with the financial implications of the services that they receive, at least at the time they receive them. In conventional full coverage health insurance, for example, the individual pays a premium and her insurance company pays for any health services that are received as and when health care is required. In many countries a form of public or social insurance exists and again individuals contribute towards the cost of the health services they receive, not at the time at which they are received, but through public or social insurance premiums or general taxation. It is conventional to refer to this second characteristic of health services as the *third party payer* principle.

The third party payer principle is found in almost all health finance systems. This prompts the question: what is the reason for its prevalence? One unavoidable conclusion is that it is preferred by individuals themselves, since both profit-seeking companies and public providers offer third party payment. Given the often high cost and intrinsic uncertainty surrounding the demand for health care, the explanation for this preference is in part the same as for any kind of insurance; that is, it offers a form of protection from the impact of unexpected large bills.

In the case of health services, however, it is likely that the explanations for third party payment go somewhat deeper. Coping with illness can be a source of considerable stress for the individual, both physical and psychological. Having a third party payer can at least shift the extra burden of financial responsibility at the time of illness. Furthermore, as we pointed out above, individuals typically have imperfect information about health services and in such circumstances they may benefit from a third party who, because they have a financial stake in the cost of meeting the individual's health care needs, has an incentive to acquire information that might enhance the quality of any decisions that need to be made. It may also be more efficient – because of, for example, *economies of scale* – for a single payer to acquire that information rather than multiple individual consumers. Large third party payers may also have additional *bargaining strength* in their dealings with health care providers and therefore be able to obtain better financial terms than those obtained by individual users themselves.

One further motivation for third party payment that may be especially relevant to health services is worth noting. In the case of health care, a desire to see services made available independently of an

individual's financial circumstances is very strong. This means that third party payment is often justified on grounds of *equity*. By using a third party payer arrangement, different forms of cross subsidisation between income groups and between the sick and those in good health can be undertaken. It is sometimes argued that efficiency and equity arguments should be considered separately. A common view among economists is that payment systems should be designed to ensure efficiency whilst income redistribution through the tax system is used to address equity concerns. The problem with this view is that most practical policies have both efficiency and equity consequences and so they need to be considered simultaneously. For example, a decision to charge for screening services may reduce utilisation, and thereby increase efficiency by deterring some marginal demand – for which the social benefit is less than the cost – but it may impact disproportionately on low income groups thereby raising equity as well as efficiency concerns. It may also be difficult, for political reasons, to pursue equity objectives through straightforward income distribution policies. Those wishing to promote egalitarian policies may therefore choose to pursue them through a 'second best' route as part of policies dealing with the provision of specific services such as health care to which access is deemed to have a social importance beyond the preferences of individuals.

We have dwelt on the reasons for third party payment at some length because cost sharing represents an erosion of this principle. Put another way, cost sharing as a policy can be seen as relaxing a constraint under which many health care systems operate. The reasons for the constraint coming into existence in the first place need to be understood if any evaluation of the consequences of relaxing it is to be undertaken. As with any constraint, the relaxation of the strict third party payment principle can be expected to confer some benefits. The problem is deciding whether these benefits outweigh the costs. A pre-requisite for this task is a framework that enables the nature of the relevant costs and benefits to be understood. In the next section we describe an economic framework which seeks to meet this requirement.

2.2 An evaluative framework for cost sharing

At least two aspects of health care systems can be discussed from an economic perspective. The first is the overall level of expenditure that is incurred by a society in delivering health services to its population. By analogy with the study of economic systems this overall measure of performance is often called macro efficiency. However, as we argue below, the concept has very little to do with efficiency as conventionally defined by economists and so we prefer to use the term *macro performance*. Second, a more detailed assessment of health systems can be obtained from considering the precise combination, quality and cost of health services that are delivered. Again by analogy with standard economic methods this can be referred to as *micro performance*. We consider each of these in turn below.

2.2.1 Macro performance

It is undoubtedly the macro performance of health systems that gives rise to most public discussion and, somewhat unfortunately, provides much of the impetus for policy debate. Measured in terms of the overall cost, almost all health care systems are becoming larger. They are consuming ever greater quantities of resources. In most countries, the growth rate exceeds that of the growth rate of the economies of which they are a part. So one of the most important stylised facts of health systems is that they are growing in real terms and relative to economies generally. Some idea of the extent of this growth, of how it compares across different economic systems and of how it relates to economic growth can be gauged from Table 1 on page 9.

As the table shows almost all OECD countries have experienced annual rates of real growth in health expenditure of four per cent or more over the period 1960-1992. For most countries these rates of growth have resulted in a steady rise in the share of GDP devoted to health. By 1992, the positions of the United States and Canada – with 14 per cent and 10.3 per cent of their GDPs devoted to health respectively – stood out as particularly high cost systems. But they were not alone in experiencing cost escalation. Between 1980 and 1992, for example, only Ireland and Sweden succeeded in reducing the share of GDP devoted to health.

While these figures are striking, their interpretation in any kind of evaluative framework is difficult (Aaron, 1996). For example, standard economic arguments would suggest that the composition of GDP will change as an economy grows reflecting the changing pattern of demand for goods and services that generally accompanies higher incomes. In other words the growing importance of health expenditures in these countries may simply be a reflection of the fact that health services are demanded more and more as individuals become richer i.e. health services are 'luxury' goods (Pauly, 1988). If this interpretation is accepted, together with the fact that technological

Table 1 **Health care expenditure as a share of GDP and annual percentage growth in real terms, OECD countries, 1960-1992**

Country	*1960*	*1970*	*1980*	*1990*	*1992*	*Annual % growth in real terms 1960-92*
Australia	4.9	5.7	7.3	8.2	8.8	4.1
Austria	4.4	5.4	7.9	8.4	8.8	3.2
Belgium	3.4	4.1	6.6	7.6	8.2	5.2
Canada	5.5	7.1	7.4	9.4	10.3	4.8
Denmark	3.6	6.1	6.8	6.3	6.5	4.7
Finland	3.9	5.7	6.5	8.0	9.4	6.2
France	4.2	5.8	7.6	8.9	9.4	6.8
Germany	4.8	5.9	8.4	8.3	8.7	4.3
Greece	2.9	4.0	4.3	5.3	5.4	6.6
Iceland	3.5	5.2	6.4	8.2	8.5	6.5
Ireland	4.0	5.6	9.2	7.0	7.1	5.5
Italy	3.6	5.2	6.9	8.1	8.5	5.9
Japan	3.0	4.6	6.6	6.6	6.9	8.5
Luxembourg	–	4.1	6.8	7.2	7.4	5.8[1]
Netherlands	3.9	6.0	8.0	8.2	8.6	4.1
New Zealand	4.3	5.2	7.2	7.3	7.5	–
Norway	3.3	5.0	6.6	7.5	8.3	5.1
Portugal	–	3.1	5.8	5.4	6.0	–
Spain	1.5	3.7	5.6	6.6	7.0	9.4
Sweden	4.7	7.2	9.4	8.6	7.9	4.1
Switzerland	3.3	5.2	7.3	8.4	9.5	4.6
Turkey	–	–	4.0	4.0	4.1	–
United Kingdom	3.9	4.5	5.8	6.2	7.1	4.0
United States	5.3	7.4	9.2	12.4	14.0	4.8

1 Data are for the period 1970-92

Source: OECD, 1996

progress in the health care tends to result in better, but more expensive, treatments (see Baumol (1995) for a discussion of the reasons for this), and account is also taken of the higher costs of treating the increasing numbers of elderly people in many countries, it becomes extremely difficult to evaluate macro performance. Indeed it is difficult to even comment meaningfully on macro performance without more information, particularly relating to micro performance. If, for example, the increase in health expenditures has been accompanied by a decline in the volume and/or quality of services that are being delivered over time, or health outcomes have deteriorated, there might be a cause for concern. But in isolation, increasing expenditure is indicative of neither success nor failure of health systems.

Increased expenditure must of course be financed. This obvious fact might explain a pre-occupation with policies designed to control the increasing level of health expenditures. In the context of private health care systems this implies that insurance premiums must rise to match increased expenditure. To the extent that premiums are in practice met by employers, and so contribute to employment costs and the price of manufactured goods, there may be a concern about competitiveness in the international marketplace.

Public health care systems also need to increase the social insurance burden on employers and employees or, if health services are financed out of general taxation, the public generally, in order to finance increased expenditures. In this last regard there has been increasing concern in recent years that tax revenues can only be increased at a real cost to economic performance because taxes, it is claimed, reduce incentives. These kinds of costs are referred to as the deadweight (welfare) burden of taxation because, proponents of this view argue, they constrain economic performance and place a burden on the economy.

Unfortunately neither the competitiveness nor deadweight burden claims are easy to evaluate. For example, the importance of competitiveness depends crucially upon what is happening in other economies.

If all economies were experiencing the same increase in employers' costs there would be no problem. Sometimes it is argued that the advanced Western economies are losing out in terms of competitiveness to the rapidly growing economies of South East Asia with their lower levels of welfare spending. However, this argument ignores the fact that it is the overall productivity of the export sector (i.e. health and non-health costs per unit output) that determines price competitiveness and not the precise composition of its costs (Fuchs, 1993). According to this view, greater productivity in the non-health sector should be able to compensate for marginally higher levels of spending on health. On the issue of deadweight welfare burden economists do not all agree as to how important an issue this is, or whether it relates more to the overall level of taxation or to the detailed design of taxation systems.

As an aside, it is also worth pointing out that recent political concern with the cost of public spending on health – and the desire to reduce the burden by substituting private for public spending – is almost invariably based upon fallacious and/or incomplete reasoning. The overall burden on the economy remains unchanged if a decrease in public spending on health is matched by an increase in private spending. Assessing the full impact of such a change upon economic performance would require complex consideration of relative productivities in the public and private sectors, the effect on savings and investment and many other factors that are rarely considered in this context. Similarly, there is a view that it is lack of public willingness to pay higher taxes that requires the substitution of private for public spending through user charges and other mechanisms. But this is not an issue of efficiency unless it is argued that expenditure is too low so denying efficient treatment to some patients. We should also note in this context the argument (Evans, Barer, and Stoddart, 1995) that top up private finance may be suggested by health care providers and suppliers in order to relieve pressures to be efficient from a cash constrained public third party payer.

Perhaps a more defensible explanation of the concern that policy makers express about macro performance is a general belief that the sort of growth that has been seen in health care expenditures can only be the result of (i) an inappropriate expansion in demand (to encompass services that are of dubious value to patients) and/or (ii) manipulation of demand by health service providers (i.e. supplier induced demand), and/or (iii) cost inflation, not related to any increase in the quality of health services that are being delivered. Put more simply, macro performance may be taken by many to be indicative of a failure of health systems at the micro level. These concerns apply both to those countries with very high levels of expenditure (e.g. the US) and those with far lower levels of spending (e.g. the UK where arguments are heard both that the NHS needs more money to improve macro performance, and that more can be achieved with existing expenditure through improved micro performance). For this reason, we now focus on micro performance.

2.2.2 Micro performance

In discussing the micro performance of the health care system, we shall distinguish between considerations of *efficiency and equity*.

Micro efficiency

Two notions of efficiency are relevant for assessing the performance of a health care system; namely, *productive* efficiency and *allocative* efficiency.

Productive efficiency is achieved when a given level of output of a good or service is provided at minimum cost. This definition corresponds to the non-economist's notion of value-for-money and implies that inputs are employed in a way that does not involve any waste or slack. Allocative efficiency is a more general concept. It encompasses productive efficiency but also measures the extent to which the overall mix of goods and services produced is consistent with consumers' preferences.

Because the concept of allocative efficiency is central to the evaluation of any economic system, it is worth dwelling on its complexities and also some of its limitations in the context of health care policy. According to the conventional definition, allocative efficiency will be improved whenever additional health services are provided that are valued by users at more than they cost to produce. Conversely, if services are provided that are valued by users at less than it costs to produce them, this is inefficient. The optimal level of production will occur when the marginal benefit obtained from the last unit of service is equal to the marginal cost of producing it. This can be called the *marginal benefit-cost* requirement.

Complications arise, however, because in the case of health care it is often the *quality* of a service that is increased rather than the quantity. Thus, for example, a more sophisticated diagnostic test or a more efficacious drug may be provided instead of a lower quality alternative. Because variations in quality form such an important dimension of health services, it is worthwhile expressing allocative efficiency in terms

that specifically address this aspect of the service. Thus an efficient quality of service implies that any improvement in quality should only take place if the benefit that it yields to the patient is greater than it costs to provide. If it does, then efficiency requires that the quality improvement is provided. We shall refer to this as the *marginal quality benefit-cost* requirement.

In our subsequent analysis of cost sharing we shall draw on these two concepts of efficiency in order to assess the performance of this particular policy instrument. However, it should be borne in mind that these concepts, as conventionally employed, do have certain limitations in the case of health care. Most notably, they are based upon a view of the well-informed, rational consumer who knows her own best interests. As we have already pointed out, the widespread existence of imperfect information in health care raises doubts about the plausibility of this assumption. Given the imperfect information possessed by patients, many decisions regarding the use of services are made by doctors on their behalf rather than by patients themselves. This calls into question the validity of a model in which the individual's assessment of likely benefits is the driving force for behind decisions about health service use. Finally, health care is an area where – because of the existence of external costs and benefits – social benefit may diverge from private benefit. In these circumstances, the efficient level of health service utilisation may be more (or less) than the level chosen by the individual on the basis of a calculation of private costs and benefits.

In our subsequent analysis, we draw on the considerable insights offered by the economists notions of allocative efficiency but are careful to draw attention to the limitations.

Equity considerations

Notions of fairness, social justice or equity in the delivery of health services are hard to define and likely to be contentious. They are nevertheless an important part of any policy discussion. Indeed the desire to promote equity in relation to health care is a fundamental principle of most health care systems, including the NHS. For this reason, the subject has attracted a good deal of attention from health economists.

In their review of the subject, Donaldson and Gerard (1993) set out a number of operational definitions of equity. In doing so, they distinguish between *horizontal* equity – which is concerned with the equal treatment of equals – and *vertical* equity – the extent to which unequals should be treated differently according to some defined criteria.

As far as horizontal equity is concerned they identify four possible criteria:

- Equal *expenditure* for equal need; e.g. equal nurse cost per bed ratios in all hospitals;
- Equal *utilisation* for equal need; e.g. equal length of stay for a given procedure;
- Equal *access* for equal need; e.g. equal waiting times for patients with similar conditions;
- Equal *health outcomes/reduced inequalities in health;* e.g. equal age and sex adjusted standardised morbidity rates across health regions and/or social classes.

On the subject of vertical equity, they offer two criteria:

- Unequal *treatment* for unequal need; e.g. more treatment for those people with serious diseases versus those with trivial ones;
- Unequal *payment* from those with unequal incomes; that is financing based upon ability to pay; e.g. progressive income tax ratios in a tax-financed system.

Clearly some of these definitions are more comprehensive than others. On the subject of horizontal equity, for example, equality of expenditure says little about the services that are actually delivered and how they affect health. Definitions based on utilisation and access come closer to measuring the health care actually received or available, but similarly neglect health outcomes. On the other hand, definitions based on equal health/reduced inequalities – while addressing the ultimate objective of health care – are far more difficult to operationalise and pursue through health resource allocation decisions, not least because they depend upon many factors outside the sphere of health care. Similar problems arise in deciding precisely how much unequal treatment of unequals is justified by the vertical equity criterion.

Clearly there are diverse views regarding what constitutes a good performance in relation to equity. We are unable to address all of these in this paper. Therefore, as a working approach, we have chosen to adopt a position that is informed by the vertical equity criteria. By taking account of the amounts that different groups contribute towards the finance of health care, and the health benefits that they receive from the system, we shall argue that any change that involves a relative deterioration in the position of those on the lowest incomes will be regarded as undesirable from an equity point of view.

2.2.3 The link between micro and macro performance

The potential for cost sharing to impact upon both micro and macro performance is perhaps best understood in the context of a specific illustrative example. Suppose we consider a simple policy change involving patients making a contribution towards the cost of a consultation with a general medical practitioner (GP) where previously no such contribution was required.

In the absence of any cost sharing a patient can be expected to visit a general practitioner if her perception of the benefit that will flow from the visit exceeds the non-monetary costs incurred. The latter will be in terms primarily of the time costs incurred by the patient. The two key features of health services that we have discussed above are clearly at work here. First, the decision must be made in the absence of complete information about the value of a consultation. Second, the visit will not, in the presence of full third party payment without cost sharing, involve any out of pocket expense. This may well mean that the benefit expected to be derived from the consultation by the patient will be less than the full cost of the consultation which will, of course, involve the doctor's time. Under these conditions there is a strong possibility that the *marginal benefit-cost* condition will be violated.

This simple example is, of course, no more than the formal expression of the common claim that – in the absence of consultation charges – GPs will be called upon to deal with some frivolous and inappropriate demands which are most unlikely to justify the full costs incurred. People putting forward this view sometimes look to cost sharing as a means for improving micro efficiency. We will return to this issue in more detail in Section 3 where some specific forms of cost sharing are considered.

In the meantime, however, it is important to point out that the above example also illustrates the dangers of rushing to easy conclusions. The lack of information on the part of patients about the value of consultations may result in some consultations being avoided which, as far as the individual's future health is concerned, are desirable. Without consulting a GP the patient may not find out whether a particular symptom is the result of a self-limiting and benign condition or is the result of a potentially serious illness requiring early treatment. In the latter case, cost sharing may actually generate inefficiency if many such valuable but incorrectly assessed consultation opportunities are deterred. Even if this is not the case, imperfect information may result in cost sharing simply having negligible effects. If, for example, naturally cautious individuals will go for consultations 'just to be safe' then cost sharing may not result in reduced consultations. In economic jargon this is saying that it might be the case that the demand for GP consultations is highly price inelastic.

These two observations, in turn, have implications for other aspects of performance. The effect, in this example, of cost sharing upon overall health expenditures appears to be straightforward: fewer consultations will result in less overall expenditure. But we have argued that the number of consultations need not be reduced. Alternatively, a reduced number of consultations may result in a delay of treatments that are effective and cheap and result instead in the need to pursue treatments that are expensive and risky in the longer term. In this case, the effect of cost sharing on overall spending, that is macro performance, could be undesirable.

2.3 Policy approaches to health system performance

Cost sharing is but one policy instrument that can be used in the quest to improve macro and micro performance. There are many other instruments. In this section we offer a brief consideration of some of these in order to clarify the context within which decisions about cost sharing have to be made.

A useful dichotomy in discussing health policy is between demand and supply. Policies that are directed primarily towards the consumers of health care can be called 'demand-side' policies whereas those that are aimed primarily at either the providers of services of the providers of insurance can be termed "supply-side'. According to this dichotomy cost sharing is a demand side policy which has its impact mostly upon the decisions of consumers of health services.

Both demand and supply-side approaches to health care have potentially important roles to play depending upon the particular aspects of performance that are of most concern. In practice, governments around the world have placed considerable emphasis on supply-side policies and so we consider some of these first.

Supply-side policies can usefully be categorised as either involving attempts to induce providers of services to improve performance through *incentives* or involving the direct *regulation* of providers.

Considerable attention has centered recently on incentives that might be built into the remuneration of health service providers. In particular it has been noted that payments for health services that are retrospective in nature provide little incentive for

providers to keep costs down and hence may result in a violation of the condition for productive efficiency outlined above. They may also lead to over provision through the encouragement of supplier-induced demand. As a result, many countries have developed prospective payment systems. These systems need to be carefully designed however if the marginal quality benefit-cost requirement for efficiency is to be maintained. These issues and how they might be resolved are discussed, in for example the case of the NHS, in Chalkley and Malcomson (1996). Further problems in the performance of health care systems can be addressed by providing incentives to providers through the use of actual or potential competition. These have, of course, been used extensively in a number of countries through the introduction of health reforms based on managed competition (Hoffmeyer and McCarthy, 1994).

Regulation of providers is often seen as a potential remedy for perceived deficiencies in the performance of health care systems. The usual motivation for regulation is that it provides a counter to the ability of providers who are not subject to competitive pressures and who might otherwise charge prices for services that are greatly in excess of costs. This would violate the marginal benefit-cost condition for efficiency discussed above. In an attempt to avoid these consequences, governments in different countries have applied various forms of regulation to the health care sector. One common approach has been to seek to restrict the supply of facilities (such as hospital beds) through planning regulations. Price controls governing the costs of treatment – such as the prospective payment system based upon diagnosis related groups (DRGs) in the United States – have also been applied widely. The pharmaceutical sector has been a popular target for such regulatory policy, involving a range of price and profit controls, limits to public re-imbursement, negative and positive lists and the promotion of generic substitutes (Burstall, 1990).

In addition to these supply-side policies there exist alternative demand-side policies to cost sharing. Probably one of the most widely discussed is the use of primary care doctors acting as 'gatekeepers' to secondary care in an attempt to reduce the extent to which inappropriately demanded health services are actually delivered to patients. Countries such as the Netherlands and the UK have highly developed systems of primary care gatekeeping with little patient direct access to non-emergency hospital services (Kirkman-Liff, 1994). Gatekeeping is also an important feature of the various models of managed care that have developed in the US in an attempt to control the growth in health care costs (Robinson and Steiner, 1997).

Numerous studies of international health policy have described how combinations of demand and supply-side instruments have been employed in different countries (see, for example, WHO (1996) for a review of current strategies in European health care systems). From the point of view of cost sharing, it is important to emphasise that individual instruments are rarely used in isolation. Thus, if it is thought that cost sharing has a part to play in reducing inappropriate use, but has certain undesirable consequences, it may be possible to combine it with other policies which address these concerns.

3 Forms of cost sharing and some implications for performance

We have defined cost sharing as the exposure of consumers of health services to out-of-pocket expenses in relation to the quantity of health care that they receive. This does not encompass all that can be or has been described as cost sharing in health services. For example, in the terminology of Rubin and Mendelson (1995), this definition involves only what they term 'direct' cost sharing. According to these authors 'indirect' cost sharing involves the total exclusion of third party payment for certain categories of expenditure (for example in some insurance contracts specific treatments such as in vitro fertilisation may be excluded). But while such exclusions involve consumers in out-of-pocket expenses, the fact that the user meets the entire bill means that they can hardly be described as involving a *sharing* of costs. For this reason, we concentrate upon direct cost sharing arrangements.

3.1 Forms of cost sharing

While there are many forms that even direct cost sharing can take, from the perspective of the incentives facing the consumer, the most important relationship is that between out- of-pocket payments and the quantity of a particular health good or service that is consumed. With this relationship in mind it is possible to classify most forms of cost sharing that occur in practice into one of the following categories.

3.1.1 Fixed payment

We will say there is a fixed payment if a consumer contributes a payment on receipt of a particular health good or service which is fixed independently of the quantity of the good or service that they consume. An example would be a charge levied for a hospital stay where the charge is independent of the number of days spent in hospital.

3.1.2 Variable payment

If the payment that the consumer makes varies with the quantity of a good or service consumed, this will be referred to as a variable payment. Variable payments can take many forms. However, remembering that it is the effect of extra units of the service consumed upon payment that is important, the characteristics of different schemes can be understood as follows. Let x be the number of units of service or good consumed (for example the number of days in hospital, the number of consultations with a general practitioner or the number of units of a medication) and c be the per unit cost of providing this good or service. The following forms of arrangement are frequently observed.

Linear pricing (LP)

The first form of variable payment allows the contribution of the consumer to be written in the form $S = p.c.x$ where S is the total payment made and p (a number greater than zero but less than one) is the fractional share to be paid by the patient. There are two practical examples of this form of cost sharing. One is referred to as *co-payment* or *user fee*, the other *co-insurance*. They differ only in the way in which the charge to the consumer is expressed and not in the way it is calculated. A user fee sets an effective price on the good or service of p.c for example, whilst co-insurance specifies a proportion (p) of the total cost of a good or service that the consumer must pay. This way of specifying payment means that the consumer faces a constant marginal price for each additional unit consumed and for this reason it is sometimes referred to as a linear pricing rule, (see for example, Tirole, 1989).

Fixed plus linear (F+L)

The simplest generalisation of a linear pricing rule involves adding a fixed element to the consumers payment. In this case the cost faced by the consumer of health services becomes $S = F + p.c.x$. This arrangement is sometimes called a *two-part tariff*. It has an important distinguishing feature that the marginal price faced by the consumer is considerably higher for the first unit consumed (for which it is F + p.c) than for subsequent units (for which it is simply p.c). This kind of arrangement applies in some countries to prescription medicines where the consumer pays a fixed component for the prescription plus an amount for each item of medicine dispensed.

Non-linear (full marginal price (FMP))

Many variations of the schemes described above can be devised but most forms of cost sharing actually involve simple combinations which involve setting p = 1 but setting some other criteria for payment. In this kind of arrangement, the consumer faces a price for marginal units of the good or service that fully reflects the cost of that good or service. For example, an insurance *deductible* or *excess* can be specified in such a way that the consumer pays the actual cost of the service up to the amount of the deductible, and hence faces the full marginal price up to this point, but then faces a zero marginal price once insurance cover cuts in.

Another variant of FMP applies in some countries where public reimbursement for the cost of pharmaceuticals is offered up to a ceiling price defined in terms of the cost of 'comparable' products. If, however, consumers want to use a more expensive product they must bear the additional costs themselves. This is known as *reference pricing*.

Non-linear (partial marginal price (PMP))

In some circumstances the consumer bears any unit cost that is over and above that allowed to the provider or supplier by the third party payer. This is often termed as *balance billing*. This arrangement means that the consumer faces a non-linear pricing rule in which the marginal price is less than cost.

3.2 The importance of the form of cost sharing

In the classification of forms of cost sharing described above, extensive use was made of the notion of the *marginal* price faced by the consumer and of the way in that this price might vary according to the payment scheme, or within a payment scheme, and according to the quantity of a good or service that the consumer receives. We will now consider exactly how these various forms of cost sharing may exert different effects and, in particular, the different implications for the performance of health delivery systems.

For illustrative purposes we will continue with our earlier example of patient consultation with a GP. The benchmark that we will use for comparison is that of no cost sharing. Consider an individual who will potentially consult a doctor a number of times, for different reasons, over a typical year. With no cost sharing the individual will attend for consultation provided that the perceived value of a visit exceeds the personal cost incurred. The total number of consultations that our individual will demand over a year will clearly be uncertain and will depend on her state of health. However it will also depend on the individual's assessments of the value of consultations and the value that she places upon her time. Some consultations, particularly those associated with worrying symptoms, will be perceived as very valuable, while others will be less so. Conceptually, we could rank consultations according to their perceived value. As we move down this ranking, consultations are of lower and lower perceived value. One way of viewing the individual's decisions over the year is that she will move down the ranking, continuing to go for consultations, until she is contemplating a consultation that is assessed to be of less value than the cost of her time. In the absence of cost sharing, this time cost – together with the costs of travelling to the doctor – are the only costs to consider.

The introduction of cost sharing can be expected to have an impact upon this individual's decisions in a number of ways. In the case of a *fixed payment* form of cost sharing for example (which we assume would involve an annual fee payable if any consultations are made but where the fee would be independent of the actual number of consultations) the only calculation that our patient needs to make is whether it is likely that the total value of all consultations that she could contemplate making in the year exceeds her value of time plus the fixed fee. Put another way, once our individual has decided that any consultation is worthwhile, the fixed payment will not have any effect on the number of consultations made. A fixed payment may therefore discourage some individuals from going to a GP, but will not affect decisions regarding the number of consultations once the decision to have at least one consultation has been made.

This can be contrasted with what happens when the patient is faced with a *variable payment*. In, for example, the case of a straightforward *linear pricing* structure (where the individual faces a fee for each consultation) the individual will cease to go for consultations when the perceived value of a consultation is less than the sum of the fee plus the value she places on her time. So whereas a fixed payment induces an 'all or nothing' response from the individual, a linear price causes the individual to be sensitive to the actual number of consultations that she makes. This is because she is facing a positive marginal price for a consultation. However with a conventional co-insurance or fixed fee linear pricing structure this marginal price does not reflect the full cost of a consultation (for example if co-insurance operates where the individual face a price of 50 per cent of the unit cost). As such, the benefit-cost condition for efficiency may still be violated.

Table 2 **Summary of the effect of different forms of cost sharing on the marginal price of a service to a consumer**

Form of cost Sharing	*Examples*	*Marginal price to consumer/patient*
Fixed payment		Zero (except for first unit)
Linear price	Co-insurance, user fee, co-payment (per unit e.g. per hospital day)	Less than full cost
Non-linear price (FMP)	Insurance deductible	Either zero or full cost
Non-linear price (PMP)	Balance billing	Either zero or less than full cost
Fixed plus linear	Prescription charges	Less than full cost (except for first unit)

Contrast this state of affairs with what would happen under a *non-linear* (FMP) arrangement, for example, where the individual's health insurance specifies a deductible or excess. There is now a range of possibilities. If the individual contemplates that her consultations will cost less than the deductible, each additional consultation is 'priced' at the full cost. There is therefore an even stronger incentive to reduce the number of consultations than in the case of co-insurance. However, if the individual contemplates that her particular health condition is likely to result in a series of consultations, the total cost of which will exceed the deductible, then we are back to the fixed payment scenario. Conditional upon the individual deciding to consult a GP, there will be no effect of this particular form of cost sharing on the number of consultations that are demanded.

Hence, from an economic perspective the effect of different forms of cost sharing upon the demand for services depends critically upon the impact that each has on the marginal price faced by a patient. These possibilities are summarised, for the most common forms of arrangement, in Table 2 above.

Of course, the principles governing the incentives of different forms of cost sharing are not unique to the case of GP consultations. As we shall see in Section 4 cost sharing arrangements have been used in a number of countries in relation to, *inter alia*, payments for pharmaceutical prescriptions, dental services, social care and for the payment for successive days of hospital in-patient stays. In all of these cases, cost sharing can be expected to exert an influence on the quantity of the services consumed.

As we argued in Section 2, however, the volume of services that is delivered is only one aspect of the performance of the health care system. What then are the expected effects of cost sharing upon other dimensions of performance? These are far less clear and are difficult to predict. However, to indicate some of the other impacts that the different forms of cost sharing that we have considered might have, consider the issue of quality of health services. In particular how might the providers of health services react to the adoption of cost sharing by third party payers?

A substantial literature has developed that considers incentives for the production of quality in a market based system, (see in particular Spence (1975)). This literature suggests that the effect of a change in consumer demand on quality may itself be difficult to predict. Most of the forms of cost sharing that we have considered result in consumers becoming more responsive to the prices of individual health goods or services. Hence, as far as the providers of those services are concerned, there is a similarity between the adoption of cost sharing on the part of payers and an increase in the extent of competition in the supply of health services – both result in greater price responsiveness of demand. But then, as Spence shows, the effect of increased competition on incentives to produce higher quality depends critically upon how consumers assess the quality of services that they receive in the first place.

It may be, for example, that providers faced with a loss of demand for their services will have an incentive to cut prices in order to restore demand but simultaneously to reduce quality in order to control their costs. This seems a particularly plausible possibility when, as in the case of health services, patients are not well informed about all aspects of quality. This lack of information, however, may lead providers to compete for patients by raising 'perceived' quality – through the provision of excessive amounts of high technology equipment and large numbers of diagnostic tests – regardless of their likely impact upon outcomes or cost effectiveness, so pushing up costs and prices. Taken overall, therefore, no general conclusions are possible regarding the impact of cost sharing on the quality of health services and it becomes important to consider the specifics of each case.

Similar considerations arise in connection with the impact of cost sharing on aggregate health expenditures (macro performance) and on the distribution of health care outcomes (equity). In order to draw any conclusions about the effect of cost sharing in these areas it is necessary to consider the empirical evidence. We shall do this in Section 5. Before then, however, we consider briefly the particular forms that cost sharing takes in practice in different countries.

4 Cost sharing in practice – international evidence

Table 3 **Cost sharing for health care services in OECD countries, October 1993**

Country	*General practitioner*	*Specialist*	*Drugs*	*Inpatient care*	*X-ray and pathology*
Australia	for 25% of bills average of $5	for 71% of bills an average of $8	maximum $11 per prescription	nil	included in specialists' bills
Austria	20% of the population pay 10% or 20%		2.50	$6	same as doctors
Canada	nil	nil	discretion of provinces	nil	nil
Denmark	nil except for under 3% of the population		0/25/50%	nil	nil
Finland	$0.17	$17	60% in excess of $8	$22	nil
Greece	nil	nil	10%/25%	*	*
Iceland	$9	$17 plus 40% of the rest of the cost	0, 12.5%, 25%	nil	$13
Italy	nil	*	0, $3 plus 50%	*	30%
Japan	Employees, 10% of all costs; dependents, 20%; self employed, 30%				
Luxembourg	5%	5%	20%	flat rate	
New Zealand	Balance billing	outpatients $3-$17	$2-$8 with stop loss	nil	outpatients $3-$17
Norway	$11	$16	25% if on blue ticket, maximum $43 per prescription	nil	X-ray $11
Portugal	*	$91-$213	0/30/60%	$30	
Sweden	$6-$19		first drug $15 then $1 each	$8	*
Switzerland	10%	10%	$7	$7	10%
Turkey	nil	nil	20%	$10 per day	nil
United States	Balance billing	20% in excess of $100 deductible	100%	$676 deductible for first 60 days	same as specialists

Note: All monetary amounts in $US

*Not specified in OECD study

Source: OECD, 1994

Cost sharing has been used in many countries in order to increase cost consciousness among consumers and to contain the growth in health care costs. Table 3 presents evidence on the different forms of cost sharing found in 17 countries included in a recent survey carried out by the Organisation for Economic Cooperation and Development (OECD, 1994).

As Table 3 shows a variety of co-payment, co-insurance and deductible cost sharing arrangements apply in different countries and in different health care sectors within these countries. In particular, it is noticeable that there are examples of both fixed and variable payment systems and, within the latter category, examples of both linear and non-linear pricing. As we argued in Section 3, these can be expected to present consumers of health care with very different incentives in relation to utilisation.

Table 4 presents international evidence on cost sharing in primary care in rather more detail by relating it to the system of payment for primary care providers found in different countries. It also presents data on annual doctor visits per capita in each country. These data are also taken from the 1994 OECD study but are

Table 4 **Cost sharing in primary care**

	Annual doctor visits per capita, 1989	*Type of payment to primary care providers*	*Cost sharing*
Australia	8.9	Fee-for-service	25% balance billing
Austria	5.8	Fee-for-service	20% of the population pay 10% or 20%
Canada	6.8	Fee-for-service	None
Denmark	5.6	28% capitation, 63% fee-for-service, 9% other	None
Finland	3.5	Salary	$0.17
Greece	5.4	Salary	None
Iceland	4.2	Salary	$9
Italy	11.0	Capitation	None
Japan	12.9	Fee-for-service	10%, 20%, 30%
Luxembourg	*	Fee-for-service	5%
New Zealand	*	Fee-for-service	Balance billing
Norway	*	35% salaries, 65% fee-for-service	30% costs for selected items
Portugal	2.8	Salary	None
Sweden	2.8	Salary	$6-$9
Switzerland	*	Fee-for-service	10% of cost
Turkey	*	Salary	None
United States	5.3	Fee-for-service	Balance billing
Belgium	7.6	Fee-for-service	Self-employed pay full cost
France	7.2	Fee-for-service	25% including balance billing
Germany	11.5	Fee-for-service	None
Ireland	6.6	Fee-for-service if higher income; capitation if lower income	None if lower income
Netherlands	5.5	Fee-for-service if higher income; capitation if lower income	None if lower income
Spain	6.2	Salary/capitation	None
United Kingdom	5.7	Capitation	None

Note: All monetary amounts in $US

*Not specified in OECD study

Source: OECD, 1994

supplemented with data for seven other countries from an earlier study (OECD, 1992).

The data presented in Table 4 confirm the importance of looking at cost sharing in the context of other policies, as emphasised earlier in this paper. They suggest that salaried or capitation payments, in comparison with fee-for-service, exert a negative effect on doctor visits (through supply-side incentives) whereas cost sharing seems to have very limited impact in this sector and at this level of analysis.

We shall return to the question of empirical evidence on the impact of cost sharing in Section 5. For the moment, however, it is worth noting that, in most of the countries covered by the OECD study, user charges have been increased to a moderate extent, and extended to more items, during the 1990s. This policy has not been applied uniformly, however. To illustrate the differential emphasis placed upon cost sharing, we provide below some brief summaries of the form that it has taken in five different countries, namely, the United States, Canada, Germany, Sweden and the United Kingdom. These countries have been selected to illustrate variations both in the extent to which health care is publicly and privately funded and in the precise form that both public and private finance takes. Apart from the OECD studies cited above, these accounts draw heavily on three main sources: Ham *et al.*, (1990); Hoffmeyer and McCarthy (1994) and Rubin and Mendelson (1995).

4.1 United States

The United States is unusual among advanced, industrial countries for its high level of dependence on private health expenditure. Approximately 60 per cent of the population is covered solely through private insurance with a further 12 per cent of the population covered through a combination of private and public insurance. Around 13 per cent of the population are covered by public insurance programmes either through the *Medicare* programme for elderly people or the *Medicaid* programme for low income groups. The form and extent of cost sharing varies considerably between the public and privately financed sectors.

Within the public *Medicare* programme cost sharing exists in the form of deductibles, co-insurance and balance billing (i.e. an extra payment levied on patients in addition to the fee received through the Medicare programme) for both hospital inpatient and outpatient services. For hospital inpatient services, the programme specifies cost sharing as follows:

- A deductible is payable, equal to the average cost of a hospital inpatient day ($696 in 1994), for up to the first 60 inpatient days;
- For the 61st to 90th days, a co-insurance payment of 25 per cent of the deductible applies ($174 per day in 1994);
- A larger co-insurance payment equal to 50 per cent of the deductible ($348 per day in 1994) applies for days beyond the 90th day;
- Full costs apply beyond 90 days when the supply of 60 lifetime reserve days has been exhausted.

For physician and some ancillary services there is a combination of a deductible ($100 in 1994) and 20 per cent co-insurance. Balance billing is also pursued, although this has been limited since 1993.

The potentially heavy burden of cost sharing in the case of long hospital stays has led to many elderly people supplementing Medicare with private insurance – the so called 'Medigap' policies. An estimated 90 per cent of elderly Medicare beneficiaries have some form of supplementary insurance. Moreover, because Medicare only reimburses for the cost of selected pharmaceuticals (such as those for patients with end stage renal failure), 65 per cent of Medicare beneficiaries also have supplementary cover for prescription medicines.

The public *Medicaid* programme for low income households is a joint federal government – state responsibility. Because states are granted considerable autonomy in the way that they run their programmes, there is a wide variation in approaches to cost sharing. Individual states are permitted to implement nominal cost sharing on all services with the exception of those provided to members of what are called 'needy populations'. This restriction prevents cost sharing being applied to services for children (under 18), services related to pregnancy or family planning. Excluding these exceptions, cost sharing varies from state to state with typical co-payments of between $2 and $5 for prescriptions and office visits being required (1992 data). Prescription drugs are the most commonly targeted programme for cost sharing, with some states only offering reimbursement for drugs in the state formulary.

Within the private sector, cost sharing has become a widespread mechanism aimed at controlling rapidly rising costs. Practically all employees enrolled in traditional, fee-for-service (FFS) plans are required to pay deductibles. These have increased substantially over the last ten years so that, in 1992, the average deductible was approximately $200.

Faced with the rising costs associated with traditional indemnity insurance, fee-for-service plans, there has been a rapid growth in managed care organisations in the US during the 1980s and 1990s. These organisations offer health care to enrollees on the basis of fixed annual subscriptions. Health maintenance organisations (HMOs) – including staff, group and network models – are the prototypical form of managed care organisation but a number of other variants have grown up alongside, and in competition, with them. These include preferred provider organisations, independent practice associations and point-of-service plans (Robinson and Steiner, 1997). In contrast to FFS plans, HMOs typically do not impose deductibles, but co-payments are common. These average between $5 and $7 for visits to a physician depending on the precise form of HMO. Other forms of managed care organisation, such as point-of-service plans and preferred provider organisations also tend to emphasise co-payments and co-insurance, rather than deductibles.

4.2 Canada

Health services in Canada are mainly publicly financed and privately provided. The whole population is included in the public health scheme, funded through taxation, and encompassing all hospital and medical services. Responsibility for financing these services is shared between the federal government and the provincial governments. Private finance is concentrated in those areas not included in the public scheme such as dental care, drugs required

out of hospital and nursing home care. In contrast to the UK, private insurers are not allowed to offer insurance for services included in the public scheme.

Canada is the only country included within the OECD study which, with the exception of pharmaceuticals, levies absolutely no co-payments in the case of health services funded by the Federal government. Although coverage varies slightly by province, in most cases public funding will cover the full costs of all medically necessary hospital services including accommodation and meals, nursing services, diagnostic procedures, operating theatre services and radiotherapy and physiotherapy services. For a number of years, doctors in some provinces were able to charge patients a fee in addition to the payment they received from the government. However, this form of balance billing – or 'extra billing' as it was known in Canada – disappeared as a result of legislation passed in 1984.

As mentioned above, private insurance is not able to be offered for services covered by the public scheme but many Canadians receive private health insurance through their employers for extras such as ambulance transport, private or semi-private hospital rooms, cosmetic surgery, dental and eye care, pharmaceuticals and medical devices. Pharmaceuticals account for between 70 and 80 per cent of total spending under private insurance.

A variety of cost sharing schemes in relation to pharmaceutical spending exist in both the public and private insurance sectors. Examples of co-payments, deductibles and co-insurance can all be found, with variations in practice between individual provinces. For example, in the province of Quebec, elderly people are entitled to a pharmaceutical benefit but subject to a co-payment of $2 (Canadian) per prescription up to an annual out-of-pocket cost of $100. In Manitoba, people under 65 are liable for 40 per cent of drug costs after an annual deductible of approximately $290.

4.3 Germany

Germany provides comprehensive health services to its population through a mixture of social insurance and private insurance. There is a compulsory social insurance scheme for people earning below a defined income level. Those with incomes above this level may opt out of this scheme and take out private insurance offering equivalent cover. In practice, around 90 per cent of the population are members of the social insurance scheme.

Recent health service reforms have extended the extent of cost sharing and, partly as a consequence of this development, there has been an increase in the number of people taking out private insurance to supplement cover provided by social insurance.

A flat co-payment per day (12 DM in 1994) is payable for hospital inpatient and rehabilitation services up to 14 days. In the case of dental treatment, patients pay between 40 and 60 per cent of the cost. It is, however, in the area of pharmaceutical expenditure that the increased use of cost sharing has been most marked.

During the 1980s German expenditure on pharmaceuticals – as a proportion of the total health budget – was approximately double the OECD average. In an effort to contain these costs a system of 'reference pricing', along with higher co-payments, was introduced. Germany's reference pricing system means that the social insurance scheme will pay the price of a drug in full if a generic brand priced at the reference price is used. However, if a patient chooses a more expensive drug instead of a generic equivalent, then the patient must bear the full cost of the difference. Until 1993, drugs not covered by the reference pricing system were liable to co-payments of between 3 DM and 7 DM depending on their price. Since 1994 the level of co-payment has remained in the range of 3-7 DM but is now related to package-size.

4.4 Sweden

Sweden provides an example of a country in which health services are overwhelmingly publicly financed – with over 90 per cent of funding coming from general taxation and social insurance – but where extensive use is made of co-payments.

Taxes are levied at the national, county and municipal levels. Around 70 per cent of public health expenditures are met from county and municipal tax revenues with the remainder deriving from social insurance and national tax revenues. Social insurance funds for the health service are raised as part of general social insurance contributions as paid by employers on behalf of their employees.

Patients make co-payments for most medical services including inpatient care, outpatient services and pharmaceuticals. In total, co-payments account for approximately 10 per cent of national health care costs. The actual level of co-payment is set by individual county councils. Typical levels of payment are: 70 Crowns per day (US$10) for inpatient treatment, 100 Crowns (US$15) for a primary care visit or an outpatient consultation, and 35 Crowns (US$5) for a visit to a district nurse. Pharmaceuticals are subject to fixed plus linear pricing whereby the first item of prescription is charged at 120 Crowns (US$17) and thereafter each additional item is charged at 10 Crowns (US$2).

4.5 United Kingdom

Public expenditure accounts for approximately 85 per cent of total expenditure on health in the UK. Health services provided by the National Health Service (NHS) are funded mainly through general taxation. The majority of services provided through the NHS are not subject to cost sharing, although patient charges do exist in relation to pharmaceuticals, ophthalmic and dental services.

Spending on pharmaceuticals represents over 10 per cent of the total NHS budget and has increased steadily over recent decades. Prescription charges have been levied almost since the inception of the NHS and currently amount to £5.65 per item (about 40 per cent of the average prescription cost) (Freemantle and Bloor, 1996). There are, however, widespread exemptions from charges for children under the age of 16, elderly people, those on low incomes and for people with specific chronic diseases. By 1995-96, 84 per cent of prescriptions were dispensed to people claiming exemption (Eversley and Webster, 1997).

From 1953 to 1969, pharmaceutical prescriptions were the largest source of NHS income from charges. Since 1969, however, they have been exceeded by dental charges. Thus in 1995, income from prescription charges was £303 million whereas income from dental charges – which represented around 30 per cent of the total cost of general dental services – amounted to £479 million (Eversley and Webster, 1997).

During the 1980s charges were also introduced for eye tests and the subsidy for NHS spectacle frames was reduced. These measures were a part of the general deregulation of ophthalmic services. In recent years there has also been more emphasis placed upon charges for superior hotel services in NHS hospitals and charging for the treatment of private patients in NHS facilities.

Despite all of these policy initiatives, however, charges remain a very small part of total NHS income. In 1995/96, for example, they amounted to only 2.4 per cent of the total (Eversley and Webster, 1997).

4.6 Conclusion

The brief overview of international experience presented in this section has indicated that cost sharing is widespread. However, the emphasis placed upon this instrument varies between countries and between health sectors within these countries. Summarising the latter evidence, the authors of a recent WHO study (WHO, 1996) claim that most western European countries place little emphasis on cost sharing in relation to physician and hospital services, but that its use is widespread in relation to pharmaceuticals. They go on to point out that in most countries co-payments for first contact, inpatient and outpatient care tend to be nominal and are often accompanied by comprehensive exemptions. But there are exceptions. OECD studies show that countries as diverse as Japan and Portugal have placed heavy emphasis on cost sharing across the board in recent years.

Our own review of the experience of five selected countries demonstrates this variety. Countries as different in terms of their socio-economic policies as the United States and Sweden make extensive use of cost sharing, whereas it is not used at all in the federally mandated and publicly funded health care sector in Canada. In Germany it has been used to address the problem of cost escalation in a particular sector (pharmaceuticals) and in a particular form (via reference pricing). A similar emphasis on cost sharing in the case of pharmaceuticals is evident in the UK. Table 5 presents our attempt to summarise this disparate experience by identifying the main forms of public sector cost sharing found in these countries in terms of the pricing structures presented in Section 3. In that section we discussed the theoretical expectations arising from the different pricing structures. In the next section we move on to consider the empirical evidence on the actual performance of cost sharing in practice.

Table 5 **Forms of cost sharing**

Country	*Types of service*		
	Inpatient	*Ambulatory*	*Pharmaceuticals*
United States	Non-linear (proportional)	Various	Various
Canada	None	None	Fixed or linear
Germany	Linear (with upper bound)	Various	Non-linear (proportional)
Sweden	Linear	Linear	Linear
United Kingdom	None	None	Fixed

5 The effects of cost sharing on performance – a review of evidence

5.1 Measuring macro and micro performance

Our discussion of the economic framework for the analysis of cost sharing suggests that it is useful to distinguish between micro and macro performance. Unfortunately, almost all studies of cost sharing have focused narrowly on the impact of cost sharing on the overall demand for services and hence macro performance. We will follow convention at this point and view any reduction in overall utilisation, and hence overall cost, as synonymous with an 'improvement' in (macro) performance, while drawing attention to the limitations of this dimension of performance as we have discussed previously. It should also be borne in mind that those people viewing cost sharing as a means of generating extra revenue for expenditure on health care would clearly define a macro 'improvement' differently; so we also comment on the revenue raising implications for the findings on utilisation.

As the discussion in Section 2 makes clear, it is not obvious how micro performance should be measured. Hence, our discussion of this must necessarily make certain assumptions. When considering the appropriate level of provision of services – which requires the benefit-cost condition to be satisfied – we will assume that any reduction in demand for medically necessary or effective services, other things remaining equal, does not contribute to micro efficiency. Such efficiency is only ensured if reductions in demand are for ineffective or inappropriate treatments, drugs or services. (There is, of course, the vexed question of determining which procedures are effective and which ineffective. As we shall see, in practice, researchers often base the decision on the views of independent medical experts).

Micro performance with respect to equity is necessarily more contentious. Following our earlier discussion, we will assume that net changes in costs and in health benefits that impact adversly upon individuals or families occupying the lower end of the income distribution are undesirable and compromise the equity aspect of performance. Even if health status is unchanged but all net incomes (i.e. after deducting out of pocket expenses on health) are reduced proportionately by cost sharing, we assume that this is to the detriment of the equity objective. In this case, cost sharing is equivalent to a regressive tax.

The issue of the quality effects of cost sharing, which we discussed under the heading of the marginal benefit – cost requirement for efficiency, has never been seriously addressed in empirical research on cost sharing. Other than suggesting ways in which this aspect of performance might be judged in the future, we will remain similarly silent on this issue.

5.2 Methodologies for assessing performance

Two different approaches are discernible in the empirical literature. The dominant approach, which characterises much empirical research in economics, is non-experimental and relies on the observation of human behaviour in uncontrolled environments. Hence, whenever the degree or form of cost sharing for certain health services is changed, demand for those services and health care outcomes might be observed and the effects of cost sharing inferred. In this approach there is always a danger that the observed behaviour or outcome has resulted from other, perhaps unobserved, influences on individual decision making. The quality of results derived by this method therefore depends crucially on the careful specification and testing of statistical or econometric models. The alternative approach is experimental, in which cost sharing is manipulated in a controlled environment and individual responses to different degrees or forms of cost sharing are observed. This approach relies on careful experimental design for its validity but produces potentially more precise and robust results. The RAND Health Insurance experiment provides a unique example of this second approach (Manning *et al.*, 1988). As such, its importance as a source of evidence on the effects of cost sharing for health services is hard to overstate and we therefore start with a review of the RAND experiment.

Table 6 **Plans included in RAND health insurance experiment**

Number of plans (category)	*Co-insurance percentage*	*Maximum expenditure of plan holder (annual)*
1 (free)	Zero	n/a
3 (25)	25%	5, 10 or 15% of income
3 (25/50)	50% for dental, mental and outpatient benefits, 25% for all other	5, 10 or 15% of income
3 (50)	50%	5, 10 or 15% of income
3 (95)	95%	5, 10 or 15% of income
1 (95/50)	0% for inpatient services, 95% for all other	$150 per person, $450 per family
1 (HMO)	0%, managed care structure	n/a

Source: Rubin and Mendelson (1995)

5.3 The RAND health insurance experiment

The RAND Heath Insurance experiment was a randomised trial of 7,708 individuals designed to ascertain the effects of cost sharing on both demand for health services and health status. The experiment was a long term one with participants recruited over a period of three years from 1974 to 1977 and followed for between 3 and 5 years. Each participant was offered one of 15 different health insurance plans, 14 of which were conventional fee-for-service plans and one of which was an HMO. The plans varied according to the extent (and more subtly the actual form) of cost sharing and also specified a maximum level of patient expenditure expressed as a percentage of income. Once this maximum had been reached no further cost sharing contribution of any kind would be required. The plans are summarised in Table 6.

The RAND researchers collected data on both usage of medical services and health care outcomes and related these to the characteristics of the different plans. This approach provides a solid statistical foundation for considering the effects of cost sharing (measured by the co-insurance percentage and maximum expenditure).

Table 7 **Annual utilisation by health plan and income group (RAND experiment)**

Health plan	*Probability of any medical use (%)*	*Probability of inpatient use (%)*	*Admissions (mean number)*	*Outpatient visits (mean number)*
		Income group: lower third		
Free	84	11	0.13	4.5
25% cs	72	10	0.13	3.5
50% cs	68	8	0.10	2.8
95% cs	58	9	0.12	2.2
IDP	65	9	0.11	2.4
		Income group: middle third		
Free	87	10	0.13	4.3
25% cs	79	8	0.09	2.7
50% cs	74	5	0.07	2.6
95% cs	70	9	0.11	2.9
IDP	72	10	0.12	2.7
		Income group: top third		
Free	89	10	0.12	4.8
25% cs	86	7	0.09	3.9
50% cs	88	8	0.11	3.7
95% cs	75	6	0.07	3.1
IDP	80	10	0.12	4.0

Note: % cs = percentage of co-insurance

IDP = Ind. Deductible Plan (95% cs for outpatient services; 0% for inpatient services)

Source: Manning *et al.*, 1988

With regard to macro performance, as Table 7 shows, the RAND experiment found that utilisation (measured in terms of probability of medical use, inpatient use, admissions and outpatient visits) decreased as the level of cost sharing increased. Thus the RAND study provides support for the view that increased cost sharing is associated with lower demand and hence a lower overall cost of health service provision.

With regard to micro performance, the results of the RAND experiment are also important. The issue of whether plan holders reduced only inappropriate demand was a much researched question in the experiment. Several different measures of appropriateness were used. On hospital admissions, two independent physician reviewers examined 1,268 hospitalisations in terms of the Appropriateness Evaluation Protocol (AEP). They concluded that inappropriate admissions represented equal proportions in both the free plan and the cost sharing plans. In relation to antibiotic prescriptions, free plan members tended to have more prescriptions, but co-insurance tended to reduce both appropriate and inappropriate prescriptions. Another test of appropriateness distinguished between four categories of effectiveness in terms of medical care. Once again, the finding was that cost sharing reduced utilisation without regard for effectiveness. In total, the overall finding was that demand for all services was reduced in line so that increased cost sharing had the same impact on the utilisation of effective as well as ineffective or medically inappropriate treatments. Health status was also measured and four out of five statistically significant differences linked cost-sharing to poorer health outcomes. As such it does not appear that cost sharing is micro efficient.

Finally with regard to micro equity, the RAND experiment provides a wealth of relevant results. As Table 7 shows, cost sharing tended to be associated with particularly marked reductions in the probability of medical use and outpatient visits on the part of lower income groups. These effects were strongest in relation to services for poor children. Moreover, although there was a significantly lower probability that a low income individual would use the health service, the average cost per service that they incurred tended to be higher. In terms of our preferred measures, it appears that increased cost sharing tended to reduce both the health system utilisation and net incomes of the lowest income groups disproportionately. As such it appears to perform poorly in terms of micro equity.

It is therefore possible to conclude that the RAND experiment provides support for the following. Cost sharing probably improves macro performance by reducing overall utilisation. Cost sharing does not necessarily improve micro efficiency because it reduces demand for appropriate and inappropriate services equally. Cost sharing impacts disproportionately on low income groups and therefore performs poorly in terms of equity.

While the RAND experiment is undoubtedly the most thorough study of cost sharing to be carried out, it does have some limitations. For example, some critics have challenged its findings on macro performance on the grounds that the study was not able to measure the impact of cost sharing on the overall cost of the health care system (Evans, Barer and Stoddart, 1995). In other words, although RAND shows that cost sharing raised revenue and reduced third party payer expenditure on the services subject to cost sharing this could still mean that, over time, providers in the cost-sharing regimes would find ways of increasing activity and so provider incomes, and hence health care expenditures. The study has also been criticised for failing to take account of the fact that many decisions on utilisation are determined by doctors rather than patients. Moreover, its focus was upon non-aged (people of 62 years of age and over were excluded) and healthy populations. As such it excluded large and vulnerable sections of the population. Finally, as we noted in Section 3, cost sharing can and does take many different forms. The RAND experiment was concerned with only variations in the simplest forms of cost sharing. Up to the limit on expenditure, the coinsurance arrangements in the various RAND plans involved linear pricing arrangements. From the perspective of economic analysis, more complicated (non-linear) forms of cost sharing have the possibility of inducing stronger responses on the part of health service users. To assess whether this stronger effect exists or not in practice would require the equivalent of the RAND experiment to be repeated for a more complex set of plans. Unfortunately no such experiment exists or is likely to be considered in the near future.

5.4 Other studies

As Section 3 indicates, cost sharing is implemented in many and various circumstances and countries. Changes in cost sharing in any of these circumstances or countries can provide some evidence as to the effects.

Rubin and Mendelson (1995) have reviewed 19 studies that examine the effect of different forms of cost sharing upon health service utilisation. These studies focus upon changes in the number of visits to the doctor, hospital utilisation (admissions, length of stay)

and pharmaceutical prescriptions. All but three of the studies focused upon US populations. The studies overwhelmingly support the RAND conclusions that cost sharing is associated with decreased utilisation in all of the areas examined.

In relation to visits to the doctor, for example, reductions in the range of 20-30 per cent in response to co-payments of different levels were not unusual. Similarly, hospital admission rates were found to fall by up to 30 per cent in response to deductibles and co-insurance. Evidence on length of stay was more ambiguous with one study showing it to be unchanged and another indicating that it actually increased. Nine studies examined the impact of cost sharing upon pharmaceutical prescriptions, including two that looked at the NHS. These showed that dramatic reductions of up to 30 per cent had been achieved in the US. In the UK, more modest reductions in the range of one to three per cent were indicated. (These studies and others are discussed in more detail in Section 5.5 which looks specifically at cost sharing in relation to pharmaceuticals in the UK.)

While the magnitude of overall effects varies from study to study, and obviously depends upon the precise form and size of cost sharing – as well as other contextual factors – the general conclusion on macro performance is that cost sharing is associated with reductions in demand for services and hence overall cost in the sectors examined.

The Rubin and Mendelson review also examines the question of micro efficiency in terms of the impact of cost sharing on inappropriate demand. They identify eight studies that examine levels of appropriate care in relation to the physician and pharmaceutical sectors. All of these relate to US populations. Once again, their findings support that of the RAND experiment; that is, cost sharing reduces both inappropriate and appropriate care.

In the physician sector, they report on studies that examined a diverse set of treatments including general physicals; cardiovascular visits; care for acute, self-limiting disease; and preventive care. In each case cost sharing was associated with a reduced level of demand. In the case of some other services, however, (i.e. cancer screening and immunisation), cost sharing did not appear to affect utilisation.

In the pharmaceutical sector, three studies show that co-payments tended to reduce the demand for both essential (defined as having important effects on morbidity or mortality) and discretionary drugs, although the fall in demand for discretionary, symptomatic or drugs of limited efficacy, tended to be larger than for essential drugs.

The Rubin and Mendelson review examined links between cost sharing and health status and found evidence that raised concerns about the impact of cost sharing on the health of unemployed and homeless people. It also looked at evidence on the micro performance of different types of cost-sharing and concluded that:

- Reactions to cost sharing appeared to be service specific. Patient initiated care (such as a visit to the general practitioner) is more readily foregone than doctor initiated care;
- Cost-sharing may discourage the use of important, cost-effective preventative services;
- Condition-specific decisions on cost-sharing are common, usually on grounds of the potential expense to the patient, or based on the 'necessity' of the treatment;
- As cost-sharing increases, its marginal impact on utilisation diminishes and so larger cost sharing payments generate more than proportionate revenue;
- If cost-sharing leads to patients taking out supplementary insurance to cover their co-payments then cost-sharing is not likely to reduce utilisation.

5.5 Cost sharing and pharmaceuticals in the UK

Charges for NHS prescription medicines were introduced in 1951 and apart from the period 1965-68 have been in existence ever since. These charges have risen steeply over time. For example, the real charge (i.e. price adjusted for general inflation) rose by nearly 300 per cent over the period 1971-93 (Morrison and Reekie, 1995). At the same time, however, the substantial growth in the numbers of people exempt from payment – on grounds of age, low income, etc – since the introduction of exemption conditions in 1968, has meant that the income from prescription charges remains a small proportion of expenditure. In 1993, for example, income from charges was only 6.3 per cent of expenditure on prescriptions (Morrison and Reekie, 1995).

Despite the existence of widespread exemptions, however, it is the *price elasticity of demand* (i.e. the percentage change in demand resulting from a given percentage change in price) that has been the focus of the limited amount of empirical research carried out in the UK. The elasticity of demand (sometimes termed utilisation) determines the change in total revenue that will result from a given change in price. As such, it is an important consideration for governments seeking to raise revenue. A convenient review of this work,

together with the results of their own work, has been reported by Hughes and McGuire (1995). (See also Huttin (1994) for a review of the most important US and UK research in this area).

The previous studies reported by Hughes and McGuire are Lavers (1983), O'Brien (1989) and Ryan and Birch (1991). Lavers studied time series evidence over the period 1967-74 and obtained utilisation (i.e. demand) elasticities in the range –0.01 to –0.02; that is, an increase in charges resulted in an almost negligible reduction in demand and hence had considerable potential to raise revenues. Using rather more sophisticated econometric techniques, Ryan and Birch reported a utilisation elasticity of –0.109. This is substantially larger than the figure obtained by Lavers but still suggests that demand is very inelastic (i.e. a ten per cent increase in price would result in an approximate one per cent reduction in demand). Finally, the study carried out by O'Brien examined the whole time period, 1969-86 and two sub-periods, 1969-1977 and 1978-1986. He obtained somewhat higher estimates of demand elasticity of –0.33 for the whole period and –0.23 (1969-1977) and –0.64 (1978-1986) for the two sub-periods. The tendency for the responsiveness of utilisation to price changes to increase (i.e. elasticity to rise) over time was also noted in another study (Morrison and Reekie, 1995) where the elasticity rose from –0.096 in 1971 to –0.175 in 1993.

In their own work Hughes and McGuire use models that are similar to those used by O'Brien and Ryan and Birch but carry out various statistical adjustments which, they claim, make their results more reliable. On the basis of these procedures, they obtain an estimated short-run price elasticity of demand (utilisation) of –0.32. In the long run it rises slightly to –0.37. Their results also suggest that short run elasticity has changed over time: in 1969 it was –0.125, in 1980 it was –0.22, in 1985 it was -0.68 and in 1991 it was –0.94. These results suggest that responsiveness to price has risen steeply over time. This means that raising prescription charges is likely to raise less revenue – but lead to greater reductions in use - than it did in the past.

By taking the estimated mean elasticity of –0.32, Hughes and McGuire calculate that the rise in prescription charges from £3.75 in 1992 to £4.25 in 1993 would have resulted in the generation of an estimated £17.3 million in extra revenue. However, as they point out this gain in revenue was associated with a decrease of 2.3 million in the number of prescriptions dispensed. The absence of data on the consequences of these reduced prescriptions – in terms of health outcomes and/or subsequent costs incurred by the NHS – makes it impossible to comment on the micro efficiency of the changes.

5.6 Drawing the evidence together

Although the body of research evidence on the impact of cost sharing is not large, a consistent pattern seems to emerge from the RAND experiment, other less rigorous studies and UK research on cost sharing in relation to pharmaceuticals. This suggests that cost sharing can be successful in raising revenue from those patients who continue to use the service. In this sense it can 'improve' macro performance by reducing costs to the third party payer. However, it is unlikely to increase micro efficiency because the reduction in utilisation caused does not necessarily result from a reduced take up of less important or less effective services. Moreover, despite the widespread existence of exemptions for poorer households, where cost sharing is applied to lower income groups, it tends to be regressive (i.e. cost sharing as a proportion of household expenditure is higher among lower income groups). In this sense, it reduces equity.

6 Summary and conclusions

This paper has sought to provide an economic perspective on the subject of cost sharing in health care. To do this, it started by highlighting the importance of the third party payer principle in health care financing and pointed out that cost sharing represents an erosion of this principle. The paper then went on to present an evaluative framework for assessing the impact of cost sharing. This framework was based upon the concepts of macro and micro performance. Macro performance refers to the widespread concern with the overall cost of, or lack of expenditure control over, the health care sector – a particular concern in many countries with public taxation or social insurance based spending on health care. From a pure economic perspective, however, it is the concept of micro performance that we argue is more important. This refers to the efficiency with which health care resources are used and also to the equity of resource allocation decisions. As part of this discussion it was pointed out that cost sharing is but one policy instrument among an array of both supply-side and demand-side instruments that can be used to improve macro and micro performance.

The presentation of the evaluative framework was followed by a discussion of the different forms that cost sharing may take. In this discussion we adopted an approach that draws heavily on economic theory in order to identify the incentive structures faced by the users of health care services under different payment systems. It will have been apparent to some readers that the traditional (neoclassical) economist's approach has some limitations in this area. For example, many decisions on health care utilisation are heavily influenced (sometimes determined) by the doctor rather than by autonomous decisions on the part of the patient, as such the normal tenets of consumer theory need to be modified. Nonetheless, we believe that there is a strong case for clarifying the nature of economic incentives – as embodied in the marginal price facing users – within a rigorous theoretical framework before considering modifications and the evidence relating to actual performance.

The presentation of the evidence on cost sharing in practice started with some summary data indicating the form and extent of its use in major OECD countries. This was followed by a brief review of the way that it is used (or not used) in five different countries; namely the United States, Canada, Sweden, Germany and the United Kingdom. These countries were selected to reflect different socio-economic systems, alternative methods or financing health care and different uses of cost sharing.

Finally we reviewed the research evidence on the impact of cost sharing using our earlier framework of macro performance (cost containment) and micro performance (efficiency and equity). In this review we paid particular attention to the RAND Health Insurance experiment as this is undoubtedly the most thorough examination of the impact of cost sharing to have been carried out to date. This study suggests that cost sharing probably improves macro performance (i.e. reduced utilisation and cost) in the sectors to which it is applied but does not necessarily improve micro efficiency (because it reduces both appropriate and inappropriate utilisation) and is likely to worsen equity through its disproportionate impact on low income users.

Other studies – including a review of cost sharing and pharmaceuticals in the UK – confirm this picture. In the case of pharmaceuticals, there is evidence that cost sharing has generated substantial extra revenue and reduced utilisation, hence reducing net expenditure by the NHS as third party payer, but its overall effects in terms of micro efficiency remain indeterminate.

These findings probably explain why most OECD countries have been cautious in their reliance on cost sharing (OECD, 1992). As our own review has shown it has been used widely but, in most countries, only at the margins. Even in the area of cost containment, it has generally been used as an adjunct to more important demand and supply-side policies (e.g. regulation of capacity, prospective payment systems and global budgets.) rather than as a substitute for them. As a means of improving micro efficiency - despite the theoretical possibilities – it seems that problems relating to patients' lack of information, together with supplier induced demand and clinical uncertainty, have seriously limited the scope for cost sharing. And, of course, there are equity concerns.

Set against these arguments there are those who claim that cost sharing could be used in the UK as a form of earmarked tax that would permit higher levels of spending on health care (Hulme, 1997). According to this view, tight political constraints on public spending prevent the provision of levels of health care that public opinion polls suggest people would like. By levying some charges, it is argued, increased revenues could be raised for extra spending on health. While there may be some substance to this argument – especially if charges were related to ability to pay – it carries the danger of deterring some appropriate use and thereby compromising efficiency. It is also not clear why an earmarked tax related to health service use (which can be labelled a 'tax on the sick') is preferable to a more broadly based earmarked tax to raise money to finance the NHS.

Taking all of the above considerations into account, it is our view that it would be unwise to undertake a major expansion of cost sharing. It may have a role to play at the margins in the discouragement of inappropriate and frivolous use and in the raising of some additional revenue. Certainly, the growing pressure for a nominal charge for GP consultations in order to discourage unnecessary demands and raise income for the NHS seems to be an expression of this view. To be efficient, however, is would be necessary to devise a system that is sensitive enough to distinguish between genuine needs and unnecessary health care demands. Precisely how this could be done is far from clear. Moreover, there is evidence of different responses to cost-sharing, depending on whether it is a patient or doctor initiated service, and of potentially adverse effects from the use of cost sharing on the take up of preventative services. Such factors would need to be taken into account in the design of any scheme. On this note, though, it is, however, worth emphasising that the existing research literature is incomplete in the sense that it does not provide a means of assessing the performance of different *forms* of cost sharing. In the light of the apparent failings of simple forms of cost sharing, the challenge facing those who favour greater use of this instrument is to design and evaluate better forms of the approach.

References

Aaron, H. (1996) 'Thinking about Health Care Finance: Some Propositions', in OECD, *Health Care Reform: The Will to Change*, Paris: OECD.

Arrow, K.J. (1963) 'Uncertainty and the Welfare Economics of Medical Care', *American Economic Review*, 53.5:941-973.

Baumol, W.J. (1995) *Health Care as a Handicraft Industry*, London: Office of Health Economics.

Burstall, M.L. (1990) *1992 and the Regulation of the Pharmaceutical Industry*, London: IEA Health and Welfare Unit.

Chalkley, M., Malcomson, J. (1996) 'Contracts for the National Health Service', *Economic Journal*, (forthcoming).

Donaldson, C., Gerard, K. (1993) *Economics of Health Care Financing: The Visible Hand*, London: Macmillan.

Evans, R., Barer, M., Stoddart, G. (1995), 'User Fees for Health Care: Why a Bad Idea Keeps Coming Back (Or, What's Health Got To Do With It?' *Canadian Journal on Aging'* 14: 360-90.

Eversley, J., Webster, C. (1997), 'Light on the Charge Brigade' *The Health Service Journal*, 107: 26-8.

Freemantle, N., Bloor, K. (1996) 'Lessons from International Experience in Controlling Pharmaceutical Expenditure', *BMJ*, 312:1469-71.

Fuchs, V.R. (1993) *The Future of Health Policy*. Cambridge, Mass: Harvard University Press.

Ham, C., Robinson, R., Benzeval, M. (1990) *Health Check: Health Care Reform in an International Context*, London: King's Fund.

Hoffmeyer, U.K., McCarthy, T.R., (1994) (eds.) *Financing Health Care*, Dordrecht: Kluwer Academic Publishers.

Hughes, D., McGuire, A. (1995) 'Patient Charges and the Utilisation of NHS Prescription Medicines: Some Estimates using a Co-integration Procedure', *Health Economics*, 4:213-220.

Hulme, G. (1997) Personal Communication.

Huttin, C. (1994) 'The Use of Prescription Charges', *Health Policy*, 27:53-73.

Kirkman-Liff, B. (1990) 'Management without Frontiers: Health System Convergence leads to Health Care Management Convergence.' *Frontiers of Health Service Management* 11.1. 3-48.

Lavers, R.J. (1983) 'The Effects of Price Changes on the Demand for and Cost of Pharmaceutical Prescriptions.' Paper presented to the International Conference on Health Economics Analysis of Medical Prescriptions. Lille, France, 14-16 September.

Manning, W.G. *et al.*, (1988) *Health Insurance and the Demand for Medical Care*, The RAND Corporation.

Mass-Collel, A., Whinston, M.D., Green, J.R. (1995) *Microeconomic Theory*, Oxford: Oxford University Press.

Morrison, G.C., Reekie, W.D. (1995) *The Paradox of the Prescription Charge: Co-payments in British Pharmaceuticals*, St Andrews: Centre for Research into Industry, Enterprise Finance and the Firm, University of St. Andrews.

O'Brien, B. (1989) 'The Effect of Patient Charges on the Utilisation of Prescription Medicines', *Journal of Health Economics*, 8:109-132.

OECD (1992) *The Reform of Health Care: A Comparative Analysis of Seven OECD Countries*, Paris: OECD.

OECD (1994) *The Reform of Health Care Systems: A Review of Seventeen OECD Countries*, Paris: OECD.

Pauly, M. (1988) 'Efficiency, Equity and Costs in the U.S. Health Care System' in Havighurst, C. *et al.*, (eds.), *American Health Care, What are the Lessons for Britain?*, London: Institute of Economic Affairs.

Robinson, R., Steiner, A. (1997) *Managed Health Care: U.S. Evidence and Lessons for the NHS*, Buckingham: Open University Press.

Rubin, R.J. Mendelson, D.N. (1995), 'A Framework for Cost Sharing Policy Analysis', in Mattison, N. (ed.) *Sharing the Costs of Health: A Multi-Country Perspective*. Basel: The Pharmaceutical Partners for Better Health.

Ryan, M., Birch, S. (1991) 'Charging for Health Care: Evidence on Utilisation of NHS Prescription Drugs', *Social Science and Medicine*, 33(6): 681-687.

Spence, A.M. (1975) 'Monopoly, Quality and Regulation', *Bell Journal of Economics* 6: 417-429.

Tirole, J. (1989) *The Theory of Industrial Organisation*, Cambridge, Mass. M.I.T. Press.

WHO (1996) *European Health Care Reforms: Analysis of Current Strategies*, Copenhagen: World Health Organisation.