Copayments and priority setting in health care: balancing equity and efficiency

Paper prepared for joint meeting of Health Economics Study Group and Collège des Economistes de la Santé, Paris, 14-16 January 2004

Peter C. Smith
Centre for Health Economics
University of York
York YO10 5DD
United Kingdom

November 2003

E-mail: pcs1@york.ac.uk
Phone: + 44 1904 321443
Fax: + 44 1904 321454

Copayments and priority setting in health care: balancing equity and efficiency

Abstract

User chargers are the major source of finance for many health care systems. However, traditional approaches to health care priority setting, such as cost-effectiveness analysis, take no account of the impact on equity and efficiency of user charges. This paper therefore develops a rudimentary model of priority setting in which the fixed health care budget can be augmented by user charges. The paper uses methods analogous to models of optimal commodity taxation to develop a set of rules for the inclusion of a health technology in the subsidized health care package, and the calculation of its associated copayment rate. The results indicate that optimal copayments depend on the cost-effectiveness of the intervention, its price elasticity of demand, the epidemiology of the associated disease, and the policy maker's attitude towards equity. The model has important implications for policy making in three domains: health care priority setting, evaluation of health care technologies, and charging policy.

Keywords: copayments, equity, priority setting, health technology assessment

JEL classification: I18

Copayments and priority setting in health care: balancing equity and efficiency

Introduction

For many people, it is a matter of straightforward humanity that no-one should be denied access to needed health care on the grounds of inability to pay. Yet it remains an inescapable fact that many health systems, particularly in low income countries, rely heavily on user charges as a source of finance, and are frequently unable to protect citizens from catastrophic health care payments (McPake, 1993; Mwabu, 2001). The World Health Organization has repeatedly drawn attention to the numerous adverse consequences of this lack of financial protection, including its macroeconomic consequences as well as its manifest direct impact on personal utility, health and fairness (World Health Organization, 2001).

The inability of many health systems to escape a heavy reliance on user charges is due to three broad types of concern. First, there are those, particularly in high income countries, who argue that user charges can moderate demand for some less essential aspects of health care (Zweifel and Manning, 2000). In the light of studies such as the RAND experiment, it is clear that charges do indeed reduce demand, but whether this leads to a welfare improvement is open to question (Newhouse, 1993). In low income countries, it is questionable whether there is widespread excess use of health care, except possibly amongst a small elite.

Second, user charges are an important source of revenue for the health system. Finance data reported by the World Health Organization indicates that, weighted for population, over 50% of worldwide expenditure on health is in the form of out-of-pocket payments, accounting for over 90% of non-governmental expenditure (World Health Organization, 2002). The figures for low income countries are even higher. Thus abandonment of user charges might, at least in the short term, have catastrophic consequences for the financing of many health systems, particularly those with very limited resources available from governments or donor agencies.

Third, whilst some form of insurance might offer a Pareto improvement over a system of direct user charges, there are in practice often no institutions in place to manage any insurance function. Worldwide, pre-payed private insurance accounts for little more than 5% of all health expenditure (World Health Organization, 2002). The widespread preference for direct charges may seem perverse, given the manifest welfare gains secured through risk pooling and insurance. However, in the absence of other sources of finance or trusted institutions to undertake the insurance function, either governmental or non-governmental, citizens must resort to direct payment mechanisms.

The pressures towards reliance on user charges were for a while given added impetus by a World Bank policy to encourage privatization and user fees, a policy stance from which it subsequently retreated (Akin*, et al.*, 1987; World Bank, 1993). Yet even where a political will exists to reduce reliance on out-of-pocket payments, there is often a perception that no feasible alternatives exist, and at least direct charging might enable policy makers to provide a breadth of subsidized health care coverage that would not otherwise be available to citizens. However, charging can lead to appalling consequences for sick people, in the form of impoverishment or denial of treatment, particularly amongst the poor. The WHO estimates that in countries such as Brazil and Vietnam in any one year up to 10% of households might be faced with 'catastrophic' health expenditures (defined as 40% of income), and that such expenditures occur disproportionately amongst the poor (Xu*, et al.*, 2003).

Policy makers in many countries therefore face an agonizing choice. They can dispense with user charges, and promote free access to health care. But the limited finance available leads either to some form of non-price rationing or to a very circumscribed package of care. Alternatively, they can introduce user charges, with all the associated harm to the poor and the sick, but thereby fund a broader package of care.

In spite of the widespread importance of this fundamental policy problem, most of the economic models used to evaluate medical technologies ignore the implications of user charges for health care priority setting. The purpose of this note is therefore to extend the traditional model of priority setting in health care to incorporate the possibility that copayments may be levied on some or all medical interventions. Conventional priority setting principles are retained: that is, the prime objective of the health system is to maximize health gain, measured for example in the form of quality adjusted life years, and possibly weighted for equity. Furthermore I assume a fixed government budget for the health care system.[1] However, this can now be augmented by user charges. The discussion first introduces a simple model of individual utility, and then examines the efficient choice of medical technologies. Equity considerations are then incorporated, and some potential refinements to the model introduced. To conclude, policy implications are discussed.

The individual utility model

The impact of user charges on utilisation can be illustrated by means of a conventional model of individual utility. Suppose all individuals have identical preferences, but differ in health H and wealth Y. A health technology i is able to raise health from h to $h + b$, where g is the health gain. A user charge p_i is proposed, equal to a proportion c_i of the costs of treatment x_i. Assume a utility function $u(H,Y)$, where $u_1>0$; $u_2>0$; $u_{11}<0$; $u_{22}<0$. Then there exists a unique critical value of wealth y_i^* such that:

$$u(h, y_i^*) = u(h+b, y_i^* - p_i),$$

where $p_i = c_i x_i$. For values of y in excess of y_i^* the patient will be better off with treatment, whilst for lower levels of y the patient prefers to decline treatment. Writing $\psi_i(c_i) = y^*(c_i x_i, h, b)$ and given a well behaved utility function $u(.)$, we can assume that

$$\frac{\partial \psi_i}{\partial c_i} > 0; \frac{\partial \psi_i}{\partial h} > 0; \frac{\partial \psi_i}{\partial b} < 0 \ .$$

Then total annual demand for intervention i can be expressed as a function of the copayment rate c_i:

$$\theta_i(c_i) = \int_{\psi_i(c_i)}^{\infty} \pi_i(y)\gamma(y)dy .$$

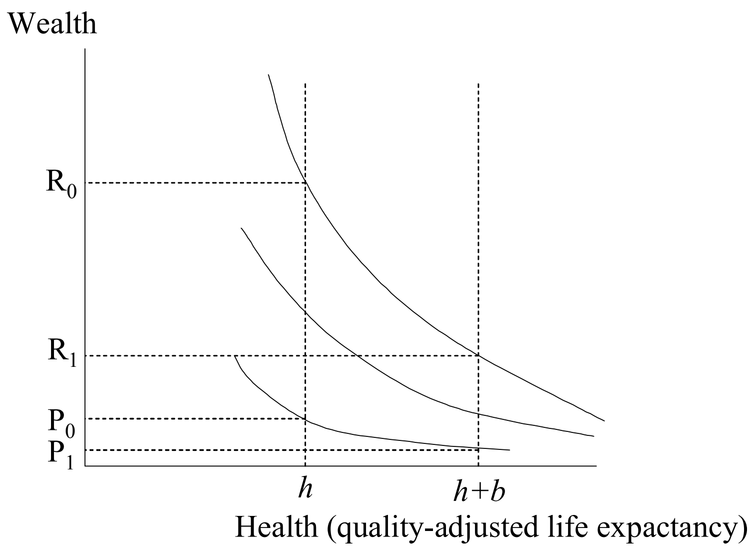
where the population is distributed according to a density function $\gamma(y)$, and the annual incidence of the condition requiring intervention is distributed as $\pi_i(y)$ in the population. Clearly $\theta_i'(c) \le 0$, and the absolute value of $\theta_i'(c)$ increases with population density $\gamma(.)$ and disease incidence $\pi_i(.)$ at the critical user charge. However, we can make no general statements about higher order derivatives. The total need for intervention *i* can be though of as

$$n_i = \theta_i(0) = \int_0^{\infty} \pi_i(y)\gamma(y)dy ,$$

the number of users if there are no charges.

To illustrate, consider in Figure 1 three individuals having identical preferences and health (health adjusted life expectancy), but differing in wealth. Each would like to be located in the top right hand of the diagram (healthy and wealthy). However, each enjoys a current health status of *h*, yielding the three prevailing indifference curves distinguished by the different levels of wealth. Suppose an intervention exists that would confer on each an improvement in health from *h* to *h*+*b*. The poorest individual has a much lower ability to pay for improved health. By undergoing the intervention, the wealthy individual would enjoy an improvement in welfare providing that the user charge was less than R_0R_1, whilst the poor individual would enjoy an improvement in welfare only if the user charge was less than P_0P_1.[2]

Figure 1: Trade off between wealth and health amongst rich and poor



Introducing heterogeneity into the patient's utility function readily yields the result that price elasticities of demand amongst the poor are higher are higher than those amongst the rich, a finding amply borne out by a wealth of empirical evidence (Gertler and Hammond, 1998). In direct terms, therefore, user charges almost always have a disproportionately adverse impact on the poor requiring a treatment for which the charge is levied. These equity considerations have led many commentators to the conclusion that user charges have no role to play in health care (Evans*, et al.*, 1995).

Note that, if the introduction of user charges permits expansion of the government-subsidized package of health care, they are effectively acting as a subsidy from one class of the sick to another. If intervention 1 is already in the subsidized package, then a new user charge for intervention 1 may make it possible to expand the package to include intervention 2. If this occurs, patients requiring intervention 2 can now receive treatment at a price below the market rate. In short, there has been a transfer from patients requiring intervention 1 to patients requiring intervention 2.

The priority setting model

Although rarely explicitly articulated, the traditional economic approach to setting health care priorities presumes a benevolent decision maker seeking to maximize health gain subject to a fixed budget constraint. The priority setting problem can be formulated as a linear programming problem. Setting aside for the moment equity considerations, externalities and any non-health outputs of the health care system, the goal is to select from all available technologies the subset that maximizes health benefits subject to the available health care budget. That is:

$$\begin{aligned} \text{Maximize: } & \sum_i \phi_i n_i b_i \\ \text{subject to: } & \sum_i \phi_i n_i x_i \leq X^* \end{aligned}$$

where $\{\phi_i\}_{i=1}^{K}$ are the decision variables, indicating the proportion of total need n_i met for intervention i, b_i are the benefits (health gains) arising from technology i for each patient treated, x_i are the associated unit costs, and X^* is the total budget available. This formulation of course assumes that interventions exhibit constant returns to scale, are independent and perfectly divisible (fractions of an intervention can be implemented). The model can nevertheless readily be augmented to accommodate relaxation of these assumptions, for example if there is a need to consider competing technologies for a particular disease. The model yields the straightforward result that policy makers should provide the set of technologies exhibiting the highest benefit/cost ratios b_i/x_i.

The inclusion of a technology in the governmental package in effect implies a 100% subsidy to all users of the intervention. In the absence of any government subsidy, interventions are assumed to be available at the price set by a competitive market. A set of government user charges therefore implies a partial subsidy for patients, so that they can use the associated technology for a price below the market rate. In what follows I assume that all patients who are willing to pay will have recourse to private health care if user charge is set equal to the market price ($c_i = 1$). I also assume the quality of care offered by the private sector is identical to that offered by the subsidized sector. So the policy problem is one of deciding whether (and how much) to subsidize specific interventions.

Introducing copayments

The charging policy problem can still be modelled as a mathematical programme, but its complexity increases considerably from the simple linear programme noted above. For each technology *i* the choice variable c_i indicates the proportion of the full market price x_i charged to users of the collective package. That is, the policy maker must choose for each technology the proportion (1-c_i) of the full market price to be subsidized. I do not consider here other forms of partial payment, such as a lump sum charge irrespective of the cost of intervention. Note that in principle c_i could be negative, implying a payment to the patient as well as free access to treatment. This possibility is considered briefly later in the paper.

As above, with copayment rate c_i, effective demand $\theta_i(c_i)$ is given by

$$\theta_i(c_i) = \int_{\psi_i(c_i)}^{\infty} \pi_i(y)\gamma(y)dy$$

I define

$$\eta_i = \frac{c_i}{\theta_i}.\frac{d\theta_i}{dc_i} = -\frac{c_i}{\theta_i}.\pi_i(\psi_i(c_i)).\gamma(\psi_i(c_i)).\psi_i'(c_i)$$

as the price elasticity of demand for the intervention *i*. Note that the elasticity will be less than or equal to zero for all levels of copayment, but that the shape of the function will be determined amongst other things by the distribution of the population and the incidence of disease within the population at the critical level of income.

If there is no concern with equity, then the priority setting problem with a fixed net public budget is to select a set of $\{c_i\}_{i=1}^{K}$ that optimizes the following programme:

$$\begin{aligned} \text{Maximize: } & \sum_i \theta_i(c_i)b_i \\ \text{subject to: } & \sum_i (1-c_i)\theta_i(c_i)x_i \le X^* \\ & c_i \in [0,1] \quad \forall i \end{aligned}$$

Providing the functions $\theta_i(.)$ are reasonably well-behaved, this problem is readily solved mathematically. The Lagrangean is defined as:

$$L(c,\lambda,\mu) = \sum_i \theta_i(c_i)b_i + \lambda\left\{X^* - \sum_i (1-c_i)\theta_i(c_i)x_i\right\} + \sum_i \mu_i(1-c_i)$$

yielding

$$\frac{\partial L}{\partial c_i} = \theta_i'(c_i)b_i + \lambda x_i\{\theta_i(c_i) - (1-c_i)\theta_i'(c_i)\} - \mu_i$$

Note that the parameter λ can be interpreted as the shadow price (expressed in terms of health) of the budget constraint. Using the Kuhn-Tucker conditions, three cases can be examined:

Case A: No user charge, $c_i^* = 0$; $\mu_i = 0$

Then $\theta_i'(0)b_i + \lambda x_i\{\theta_i(0) - \theta_i'(0)\} \le 0$

In this case $\frac{b_i}{x_i} \ge \lim_{c\to 0} \lambda.\left(1 - \frac{c}{\eta_i(c)}\right)$ and the intervention is cost-effective without user charges. Note that as the absolute value of the elasticity declines, the required hurdle for zero charges increases, as the health benefits lost through a charge are outweighed by the associated revenue. For very high elasticities, the minimum hurdle is λ.

Case B: Full user charge, $c_i^* = 1$**;** $\mu_i \ge 0$

Then $\theta_i'(1)b_i + \lambda x_i\theta_i(1) \le 0$

In this case $\frac{b_i}{x_i} \le -\lim_{c\to 1} \lambda\frac{1}{\eta_i(c)}$ and the intervention is not cost-effective with any level of user charges.

Case C: Intermediate user charge, $1 > c_i^* > 0$**;** $\mu_i = 0$

Then neither boundary condition is binding and

$$\theta_i'(c_i)b_i + \lambda x_i\{\theta_i(c_i) - (1-c_i)\theta_i'(c_i)\} = 0$$

So

$$\frac{b_i}{x_i} = -\frac{\lambda\{\theta_i(c_i) - (1-c_i)\theta_i'(c_i)\}}{\theta_i'(c_i)} = \lambda\left[1 - c_i^*(1+\eta_i^{-1})\right]$$

and the elasticity at the optimal user charge is given by

$$\eta_i^* = \frac{-\lambda x_i c_i^*}{b_i - \lambda x_i(1-c_i^*)} .$$

The top line of this expression indicates the user fee generated by the marginal patient, while the bottom line indicates the net social benefit of treating the patient. Thus the benefit-cost ratio of the technology is inversely related to the price elasticity of demand. It is important to note that the functions $\theta_i(c_i)$ are not in general concave, so the optimum described may not be unique. The results merely describe the characteristics of an optimal solution.

The results imply that an optimal solution has the following characteristics:

- For sufficiently high values of the benefit:cost ratio b_i/x_i no zero user charge will be levied, as the benefits of free access outweigh any foregone revenue;
- For sufficiently low values of b_i/x_i the full market price copayment will be levied, as the benefits of the technology are never sufficient to justify any user charge subsidy;
- For moderate levels of b_i/x_i an intermediate copayment may be levied, as the benefits of user charge revenue compensate for some reduction in utilisation;
- Where an intermediate copayment is levied, it will increase as elasticity approaches zero (when increased revenue outweighs reduced benefits);
- Where an intermediate copayment is levied, it will be decreasing in b_i/x_i.

This analysis reinforces the view that – when resources are limited – free or low price access should be targeted at conditions for which user charges would at the margin deter significant numbers of patients from seeking care. These are likely to be high cost conditions with high incidence amongst the poor.

Introducing equity

In order to introduce equity into the analysis, I assume that there is some dimension of ‘need’ along which society’s valuation of a unit of health gain increases. For this analysis, I assume that need is reflected in the measure of wealth y. There therefore exists a function $w(y) \geq 0$ that indicates the societal weight attached to a unit of health gain for a person with wealth y, where $w'(y) \leq 0$. Thus, more needy groups are more readily deterred by copayments from making use of health care, but their benefits are more heavily weighted...

The policy problem now becomes one of maximizing weighted health benefits within the budget constraint. That is, selecting a set of $\{c_i\}$ so as to:

$$\text{Maximize}: \sum_i b_i \int_{\psi_i(c_i)}^{\infty} w(y)\pi_i(y)\gamma(y)dy$$

$$\text{subject to}: \sum_i (1-c_i)x_i \int_{\psi_i(c_i)}^{\infty} \pi_i(y)\gamma(y)dy \le X^*$$

$$c_i \in [0,1] \quad \forall i$$

Differentiating the Lagrangean gives rise to the optimality conditions for interior solutions

$$b_i\psi_i'(c_i^*)\left[w(y)\pi_i(y)\gamma(y)\right]_{y=\psi_i(c_i^*)} + \lambda x_i\left\{\theta_i(c_i^*) - (1-c_i^*)\psi_i'(c_i^*)\left[\pi_i(y)\gamma(y)\right]_{y=\psi_i(c_i^*)}\right\} = 0$$

These yields the results

$$\eta_i^* = \frac{-\lambda x_i c_i^*}{b_i.w(\psi_i(c_i^*)) - \lambda x_i(1-c_i^*)}$$

That is, compared with the efficient solution, the benefit of treatment b_i is scaled in proportion to the social weight attached to the marginal patient deterred. The optimal user charge is negatively related to the social weight attached to the marginal patient. The policy message is therefore clear. If there is a concern with equity, compared with the efficient solution, charges will be reduced on interventions that deter relatively 'needy' people (that is, for which $\psi_i(c_i^*)$ is low). Zero user charges will now be more readily applied to interventions for which any fees deter low income patients, most likely high cost, high benefit interventions that threaten catastrophic impoverishment.

User charge exemptions

A further instrument in principle available to policy makers is the use of a waiver (or exemption) for high needs patients. In practice, waivers have proved difficult to implement in many low income countries, as identifying whether an individual meets the qualifying criterion is not straightforward, and providers are reluctant to enforce charges if some element of discretion exists (Bitrán and Giedion, 2002). Waivers are nevertheless easy to incorporate into the mathematical formulation. For each intervention i define a threshold level d_i of income y below which no fees are charged.

Because the social welfare function used here is concerned only with maximizing (equity weighted) health, the optimal level of d_i will never exceed the threshold charge $\psi_i(c_i^*)$ below which demand is suppressed. (If d_i were to exceed $\psi_i(c_i^*)$ there would be an 'unproductive' flow of funds out of the health system to patients who would have been willing to pay for the intervention.) The mathematical programme is therefore reformulated as follows:

$$\begin{aligned}
\text{Maximize: } & \sum_i b_i \left\{ \int_0^{d_i} w(y)\pi_i(y)\gamma(y)dy + \int_{\psi_i(c_i)}^{\infty} w(y)\pi_i(y)\gamma(y)dy \right\} \\
\text{subject to: } & \sum_i x_i \left\{ \int_0^{d_i} \pi_i(y)\gamma(y)dy + (1-c_i) \int_{\psi_i(c_i)}^{\infty} \pi_i(y)\gamma(y)dy \right\} \le X^* \\
& d_i \le \psi_i(c_i) \quad \forall i \\
& c_i \in [0,1] \quad \forall i
\end{aligned}$$

For interior solutions when the subsidiary constraints are not binding ($d_i < \psi_i(c_i)$), differentiating the Lagrangean with respect to c_i gives an identical set of equations to the situation where there are no exemptions. However, the exempt patients will effectively reduce the budget available for all (increase λ), so all copayments will (if anything) increase.

Differentiating the Lagrangean with respect to d_i gives

$$b_i\left[w(s)\pi_i(s)\gamma(s)\right]_{s=d_i^*} - \lambda x_i\left[\pi_i(s)\gamma(s)\right]_{s=d_i^*} + \sigma_i = 0$$

where the last parameter represents the Lagrange multiplier associated with the constraint $d_i \le \psi_i(c_i)$. If this constraint is not binding ($\sigma_i = 0$), this set of conditions implies that $w(d_i^*) = \frac{\lambda x_i}{b_i}$. That is, the weighted benefits to the marginal exempt patient equal the social costs, and therefore the social weight attached to the marginal exempt patient equals the inverse of the benefit/cost ratio. So other things equal the threshold for exemptions will increase as the benefit/cost ratio increases.

If $d_i^* = \psi_i(c_i^*)$, then $\sigma_i \neq 0$ and the social weight attached to the marginal patient may increase. Note that

$$\frac{\sigma_i}{[\pi_i(y)\gamma(y)]_{y=d_i^*}} = \lambda x_i - b_i[w(y)]_{y=d_i^*}$$

Furthermore, the set of first order conditions with respect to c_i becomes:

$$b_i\psi_i'(c_i^*)[w(y)\pi_i(y)\gamma(y)]_{y=\psi_i(c_i^*)} + \lambda x_i\{\theta_i(c_i^*) - (1-c_i^*)\psi_i'(c_i^*)[\pi_i(y)\gamma(y)]_{y=\psi_i(c_i^*)}\}$$
$$+\sigma_i\psi_i'(c_i^*) = 0$$

leading to the result that the marginal elasticity becomes:

$$\eta_i^* = \frac{-\lambda x_i c_i^*}{b_i.w(\psi_i(c_i^*)) - \lambda x_i(1-c_i^*) - \sigma_i/[\pi_i(y)\gamma(y)]_{y=d_i^*}} = -1$$

At the income level where the exemption ends and user charges begin, the elasticity is -1.

Of course it may be considered unacceptable to create a 'care gap' in the situation in which d_i^* is strictly less than $\psi_i(c_i^*)$, when those with needs s between d_i and $\psi_i(c_i^*)$ will not secure treatment. The mathematical programme is readily amended to rule out such possibilities by requiring that $d_i = \psi_i(c_i)$ for all *i*.

The policy maker may also be constrained to setting a single exemption income d^* for all interventions. This leads to the result that the weight attached to the marginally exempt patient must be proportional to the *aggregate* cost/benefit ratio:

$$w(d^*) = \lambda \frac{\sum_i \pi_i(d^*)x_i}{\sum_i \pi_i(d^*)b_i}.$$

Of course such a solution may for some treatments lead either to a care gap ($d^* < \psi_i(c_i^*)$), or to a situation where some patients who are willing to pay are granted exemption ($d^* > \psi_i(c_i^*)$).

Subsidizing access

I have so far considered only the possibility of user charges lying in the interval [0,1]. Yet in practice, most health systems find that many disadvantaged people are inhibited from gaining access to care, even in the absence of direct user charges (Khan*, et al.*, 2002). This might be because there are other indirect costs of gaining access (such as taking time off work, securing child care, or transportation costs), or because providers fail to provide care to disadvantaged groups because unit costs are higher. In either case, a further subsidy may be required if the 'needy' are to gain access. Such subsidy, whether paid to the patient or the provider, can be modelled as a negative user charge. Subsidies to secure access can therefore be readily incorporated into the analysis by relaxing the lower bound of zero imposed on the user charge, effectively allowing 'negative copayments'. Clearly such subsidies will reduce the size of the budget available for other interventions, and so will be targeted at interventions with high benefit-cost ratios and high elasticities at zero price.

Discussion

Many studies of health care in low income countries report low price elasticities of demand (see for example (Akin*, et al.*, 1985; Gertler and Van der Gaag, 1990)). This has led some policy advisors to infer that increased user charges are an efficient way of increasing the size of the health sector. However, such arguments ignore the manifest equity implications of a reliance on user charges, and the practical difficulties associated with administering exemption schemes. Equally, the advocates of increased fairness in health care sometimes ignore the efficiency sacrifices that arise in the pursuit of equity. The model presented in this paper seeks to offer an analytic framework that integrates equity and efficiency concerns into the priority setting process, and that can serve as a basis for debates about charging policy.

With adequate funds (in the absence of moral hazard) the optimal solution may well be to remove fees for most mainstream interventions, and thereby effectively offer universal insurance coverage for a core package of care. This obviates the need to 'tax the sick' in the form of user charges, and arrangements for premium contributions can then be designed independently of incidence of disease. Even with limited funds, the establishment of a reliable insurance function for a limited package of care is one way of mitigating the worst effects of a charging regime. The situation modelled here is therefore very much one of second best, in which insurance arrangements cannot be made widely available. The results confirm the view that – in these circumstances – government subsidies should, other things being equal, be directed at interventions that (a) have high benefit-cost ratios (b) have high price elasticities of demand, particularly amongst the poor and (c) have relatively high incidence amongst the poor. These will typically be cost-effective treatments that would otherwise lead to widespread catastrophic impoverishment.

One of the implications of the analysis presented here is that optimal solutions will be highly dependent on a nation's budget constraint, epidemiology and policy preferences. In the presence of user charges, it is highly unlikely that rankings of health technologies will be invariant to such local circumstances. The analysis therefore suggests that universal measures of the cost-effectiveness of interventions should represent just one of a number of inputs to the national priority-setting process.

The model presented here is highly stylized, but embraces most of the essential features of practical policy making in three domains: health care priority setting, evaluation of health care technologies, and charging policy. They seek to extend the principles of optimal commodity taxation into the context of health care priority setting (Sandmo, 1976). Future developments to the models could include: the introduction of quality differences between private (market) health care and subsidized health care; generalization of the social welfare function to other goods and services (effectively introducing a flexible health care budget constraint); introduction of externalities; and incorporation of economies of scope and scale. In order to keep the models as simple as possible, none of these developments has been reported here. However, depending on the policy preoccupations, each can readily be incorporated into this framework.

References

Akin, J., Griffin, D., Guilkey, B. and Popkin, B. (1985), *The demand for primary health services in the Third World*, Totowa NJ: Littlefield Adams.

Akin, J., Birsdall, N. and de Ferranti, D. (1987), *Financing Health Services in Developing Countries: An Agenda for Reform*, Washington DC: World Bank.

Bitrán, R. and Giedion, U. (2002), *Waivers and exemptions for health services in developing countries*, Washington DC: World Bank.

Evans, R., Barer, M. and Stoddart, G. (1995), *User Fees for Health Care: Why a Bad Idea Keeps Coming Back*, Vancouver: University of British Columbia, Centre for Health Services and Policy Research.

Gertler, P. and Van der Gaag, J. (1990), *The willingness to pay for medical care in developing countries*, Baltimore: Johns Hopkins University Press.

Gertler, P. and Hammond, J. (1998), "Strategies for pricing publicly provided health services", in G. Scheiber (ed), *Innovations in health care finance*, Washington DC: World Bank.

Khan, M., Walley, J., Witter, S., Imran, A. and Safdar, N. (2002), "Costs and cost-effectiveness of different DOT strategies for the treatment of tuberculosis in Pakistan", *Health Policy and Planning*, 17(2), 178-186.

McPake, B. (1993), "User charges for health services in developing countries: a review of the economic literature", *Social Science & Medicine*, 36(11), 1397-1405.

Mwabu, G. (2001), "User charges for health care: a review of the underlying theory and assumptions", in G. Mwabu, C. Ugaz and G. White (ed), *Social Provision in Low-Income Countries - New Patterns and Emerging Trends*, Oxford: Oxford University Press.

Newhouse, J. (1993), *Free for all? Lessons from the RAND health insurance experiment*, Cambridge MA: Harvard University Press.

Sandmo, A. (1976), "Optimal taxation: an introduction to the literature", *Journal of Public Economics*, 6(1), 37-54.

World Bank (1993), *World Development Report 1993: Investing in Health*, New York: Oxford University Press.

World Health Organization (2001), *Macroeconomics and Health: Investing in Health for Economic Development. Report of the Commission on Macroeconomics and Health*, Geneva: World Health Organization.

World Health Organization (2002), *World Health Report 2002: reducing risks, promoting healthy life*, Geneva: World Health Organization.

Xu, K., Evans, D., Kawabata, K., Zeramdini, R., Klavus, J. and Murray, C. (2003), "Household catastrophic health expenditure: a multicountry analysis", *The Lancet*, 362, 111-117.

Zweifel, P. and Manning, W. (2000), "Moral hazard and consumer incentives in health care", in J. P. Newhouse and A. J. Culyer (ed), *Handbook of health economics*, Amsterdam: Elsevier.

Endnotes

[1] The policy maker's broader objective should be choose a health budget so as to maximize general social welfare. Although this more general formulation would be interesting, the purpose of this paper is specifically to examine the implications of introducing user charges into the traditional health economist's priority-setting model, which assumes a fixed budget constraint.

[2] Assuming diminishing marginal utility of wealth, even when the poor are willing to pay for an intervention, the change in utility derived from securing treatment will be less than the change in utility for their richer counterparts.