Towards an index of health coverage

Discussion paper 2012/11

December 2012
Towards an index of health coverage

Rodrigo Moreno-Serra\textsuperscript{a} and Peter C. Smith\textsuperscript{b}

Imperial College Business School and Centre for Health Policy

December 2012

Contact information:
\textsuperscript{a} Email: r.moreno-serra@imperial.ac.uk (corresponding author)
\textsuperscript{b} Email: peter.smith@imperial.ac.uk

Abstract: The concept of “health coverage” is a central feature of current debates about health systems, referring in broad terms to the extent to which people can secure access to needed health services without financial impoverishment. There is growing interest in developing consistent indices of coverage in order to compare different health systems, and to track progress over time. Fundamental questions remain, however, about how to formally define and measure health coverage. This paper proposes an economic model that reconciles the various concepts of health coverage found in the literature, and can serve as a basis to develop an operational index of coverage that reflects concerns with both unmet treatment need and financial protection. The proposed conceptualization can help in the development of policy also by providing a useful framework, based on standard economic theory, to link measured health coverage to statements about social welfare and optimal policy.

Keywords: Health coverage; financial protection; access; user charges

JEL classification: I18

Acknowledgements: This paper is based on a report prepared for the World Health Organization, Geneva. We are grateful to David Evans, Priyanka Saksena and Ke Xu at WHO for their guidance. We should also like to thank Marianne El-Khoury, Laurel Hatt, Owen O’Donnell, Eddy van Doorslaer, Adam Wagstaff, Hong Wang and other participants at a workshop held to discuss the paper.
Introduction

The concept of “health coverage” is a central feature of current debates about health systems, referring in broad terms to the extent to which people can secure access to needed health services without financial impoverishment (World Health Organization 2010a; International Labour Organization 2008). The effectiveness of prevailing health insurance mechanisms, and the potential benefits of expanding access to care and financial protection, especially for poorer population groups, are not solely a concern of policymakers in low and middle income countries. A prime example of this is the heated debate surrounding the societal value of expansions of the state-run U.S. Medicaid programs to cover larger numbers of low-income citizens under the Affordable Care Act (Frakt et al. 2011; Gottlieb 2011; Sommers et al. 2012).

Globally, the interest in conceptualizing and measuring health coverage was given strong impetus by the publication of the World Health Report 2010, with the title Health Systems Financing: the Path to Universal Coverage (World Health Organization 2010a). The report focused on the role of pooled funds for healthcare, and the extent to which these were deployed effectively so as to secure access to needed health services for all citizens without financial hardship, thereby contributing towards a concept of “universal health coverage” (cf. e.g., The Lancet 2010; World Health Organization 2011). Such pooled funds usually correspond in practice to government health expenditure (and we refer to them as such throughout this paper), although there is no requirement for pooled funds necessarily to be administered by a government. Essential features of the pooling scheme are that financial contributions are unrelated to healthcare needs, and that access to a specified set of health services is thereby guaranteed to the population considered members of the scheme. The role of pooled funds in such a context is to subsidize, either partially or in full, the cost of healthcare treatments. If there is no subsidy from the pooled funds, those in need of a given treatment who are willing and able can secure access by paying its full market price. In general, those on lower incomes may be deterred from seeking treatment in such circumstances. If the government subsidizes the treatment, so that it becomes available free or below the market price to the user, potential access to treatment is opened up to larger numbers of people on lower incomes who would otherwise not be able to receive it.

The international policy debate on health coverage is usually framed around three fundamental dimensions of coverage: the proportion of the population covered; the range of needed services covered; and the proportion of costs covered by the pooled funds, as in Figure 1. A policy implication of this framework is that improvements in coverage within a health system can be secured by, inter alia:

- extending the population covered by pooled funds;
- extending the range of services covered by pooled funds; and
- reducing the proportion of direct treatment costs borne by pool members.

<Figure 1 about here>

Reflecting the high profile currently given to the topic in multilateral discussions, policymakers are increasingly committing to some notion of universal health coverage as a fundamental goal of health system design. In addition to the proposed U.S. Medicaid policies mentioned previously, important examples of related reform implementation include countries as varied as Thailand, Mexico, Indonesia, Ghana and Kenya, to name but a few (Gruber et al. 2012; King et al. 2009; Lagomarsino et al. 2012).
There is, however, no generally accepted theoretical framework within which the international policy debate on health coverage has taken place. Commentators have tended to adopt two distinct perspectives on the issue. The first, dominant in the public health literature, examines the observed utilization of health services in relation to some norm of utilization that might be expected if there were no financial or other constraints to access (Oliver and Mossialos 2004; World Health Organization 2010b). The second perspective, more commonly found in the economics literature, examines the extent to which citizens suffer financial hardship or bankruptcy as a result of using specified health services, and the potential welfare effects of interventions such as public insurance (see e.g., Wagstaff and Yu 2007; Gross and Notowidigdo 2011; Einav et al. 2010). These alternative traditions require different analytic conventions, employ different (and sometimes ad-hoc) metrics, and often give rise to conflicting and incomplete perspectives on the ultimate policy concern.

Fundamental questions remain about how to formally define and measure health coverage, as well as how to use scarce government resources optimally so as to maximize coverage. The aims of this paper are to propose an economic model that reconciles the various concepts of health coverage found in the literature, and can serve as a basis to develop an operational index of coverage that reflects concerns with both unmet treatment need and financial impoverishment. We argue that the proposed conceptualization can help in the development of policy also by providing a useful framework, based on standard economic theory, to link measured health coverage to statements about social welfare and optimal policy.

The organization of the paper is as follows. The next section gives a brief survey of the relevant available literature. We then propose a simple utility-based model of coverage, with arguments in health and wealth, which can act as a framework for analyzing the concept. We use this framework to assess the usefulness of existing metrics of coverage, and to derive a theoretical index of coverage which incorporates access to care and financial protection dimensions. In the Appendix, we also use the proposed framework to develop a model of optimal coverage design for a fixed budget constraint. The paper ends with a discussion of the implications for the design and measurement of coverage, and the associated policy implications.

The concept of health coverage: definition and measurement

In the empirical literature, and especially in the public health field, the most common approach to defining and measuring health coverage levels has been to focus on particular health services, aimed at specific target groups (e.g., vaccination of children) (UNICEF 2009). It is assumed that the chosen services are needed by anyone in the specified groups (defined by age or gender, for example), and coverage is then made synonymous with the proportion of the relevant population receiving the intervention. Frequently, researchers analyze a few services in isolation, with no explicit theoretical framework being adopted to guide the choice of health coverage metrics.

There is a particular lack of theoretical conceptualization from a micro-economic perspective. Much of the related theoretical research has focused instead on two specific health insurance aspects, predominantly within the U.S. context: the rationale for government intervention in private insurance markets and (to a lesser extent) the optimal choice of government policy instruments (see Chetty and Finkelstein 2012 for a review). The first strand of literature has examined mainly the role of health market failures as a justification for welfare-enhancing government interventions, sometimes within
richer versions of more standard economic models in the vein of Rothschild and Stiglitz (1976). The second strand of research has investigated the social welfare consequences of alternative government policies to expand insurance coverage, such as direct public provision and mandates (Einav et al. 2010, 2011; Gross and Notowidigdo 2011). In this literature, the focus is often on the conventionally defined adverse selection and moral hazard costs of higher utilization resulting from extended insurance coverage. Extended coverage in this context refers to the extra numbers of individuals who take up health insurance, or benefit from reduced cost-sharing for certain medical services. No specific attention is usually paid in such models to the value of improvements in health status, for example, generated by increased use of services by people who were previously deterred from accessing needed medical care due to financial barriers.

Outside the U.S. context, there have been a few attempts to derive metrics of health coverage from more formal analytical models. These studies seek to develop multidimensional coverage indices useful to track health system performance over time and for international comparisons, much in the spirit of the far richer literature on poverty measurement (see e.g., Alkire and Foster 2011; Atkinson 2003). However, the proposed health coverage metrics invariably lack a clear grounding in standard economic theory.

For example, an early study by Tanahashi (1978) proposes that health service coverage should be expressed as the extent of interaction between the service and the population groups for whom it is intended. The author defines five coverage elements that must be examined for a given service: (1) availability coverage, or the extent to which a service is in principle made physically available to the population; (2) accessibility coverage, or the extent to which the population can secure access to an available service, given geographical and other physical constraints; (3) acceptability coverage, or the extent to which the population makes use of an accessible service, given financial, cultural, and other barriers; (4) contact coverage, or the extent to which the provision of an acceptable service leads to contact between provider and patient; and (5) effectiveness coverage, or the extent to which contact with a provider leads to delivery of an effective service. Dimensions such as “acceptability” and “contact” can readily be given an economic interpretation, insofar as they reflect an underlying model of individual choice subject to relevant constraints. However, Tanahashi’s categorization is not based on any formal economic framework that could be used to guide the construction of operational metrics of service coverage, or to assess coverage levels for broader service packages through a unitary index.

A more formal model of “effective coverage” was proposed by Shengelia et al. (2005) and has been applied in a few empirical studies (Lozano et al., 2006; de Looper and Lafortune, 2009). They adopt an ex-ante view and define effective coverage as the fraction of maximum possible health gain an individual with a healthcare need can expect to receive from the health system. Formally, the suggested framework can be expressed as \( EC_{ij} = Q_{ij}U_{ij} | N_{ij} = 1 \), where \( EC_{ij} \) is effective coverage for intervention \( j \) for individual \( i \); \( Q_{ij} \) is the expected quality of the intervention; \( U_{ij} \) is the probability that the intervention will be received by the individual from any of the available providers; and \( N_{ij} \) is an indicator for individual’s \( i \) need for intervention \( j \). The authors then formalize a service’s expected effective coverage as:

\[
EC_j = \frac{\sum_{i=1}^{n} EC_{ij} H_{ij} \Pr(N_{ij} = 1)}{\sum_{i=1}^{n} H_{ij} \Pr(N_{ij} = 1)},
\]
where \( HG_j \) is the health gain from the service for individual \( i \). The overall health system’s effective coverage will then be the fraction of the maximum health gain that the system can be expected to deliver (between all available providers \( k \)), given the set of potential needs the population faces:

\[
EC = \frac{\sum_{j=1}^{n} EC_j \cdot HG_j \mid P_k = P_k^{\text{max}}, \forall k = 1\ldots n}{\sum_{j=1}^{n} HG_j \mid P_k = P_k^{\text{max}}, \forall k = 1\ldots n}.
\]

Shengelia et al. (2005) state that individual demand for healthcare in their framework is derived from a static, two-period utility maximization model, which is a function of individual and provider characteristics: income net of healthcare consumption, distance to the chosen provider, and responsiveness of the chosen provider. Yet the authors do not fully develop their model to specify key elements, such as the precise nature of individual preferences governing care seeking in the model and the role of service prices. There is no direct link between a utility-based social welfare function and their suggested measure of effective coverage in a health system.

In addition to the health coverage dimensions explored in the studies above, the concept of coverage has evolved to embrace the notion of financial risk protection – the idea that access to necessary health services should not cause financial hardship to users through excessive health payments (World Health Organization 2010a). Economic theory has long acknowledged the value attributed by individuals to protection mechanisms against substantial financial losses (see e.g., Knight 1921), yet the issue of financial protection has been largely neglected within the theoretical literature on health coverage discussed above. Instead, the topic has been addressed chiefly in an extensive empirical literature, where the extent of financial protection is usually measured by the relative participation of out-of-pocket payments in the total financing of the health system or, alternatively, the incidence of catastrophic healthcare payments in a population (e.g., Wagstaff and van Doorslaer 2003; Xu et al. 2003). Catastrophic spending is usually defined as occurring if a household’s health expenditures exceed a pre-defined threshold, in terms of a fixed amount or share, of a chosen income aggregate serving as living standard measure. The selection of the actual threshold above which health payments are defined as catastrophic is unavoidably arbitrary and ultimately a normative choice, with most studies considering a household incurring catastrophic health spending if the latter represents anything between 10-40% of the household’s pre-payment, non-subsistence income (see, among others, Wagstaff and van Doorslaer 2003; van Doorslaer et al. 2007; Xu et al. 2007). Some of these empirical studies also assess financial protection by looking at the extent to which medical expenses cause households to fall below a given income poverty line (impoverying health spending). Once again, the pre-defined poverty line threshold remains a normative issue.

Although not explicitly informed by micro-economic theory, such studies usually (explicitly or implicitly) invoke certain economic assumptions. One such assumption is that healthcare spending affects individual welfare by imposing an opportunity cost in terms of foregone consumption of other goods and services. This opportunity cost – and any associated utility loss – will tend to increase according to the complexity of the demanded health services and may become “unacceptably” high. This reflects a fundamental concern about the potentially high cost of healthcare, both in absolute and relative terms, as even low-cost health services may lead to major utility losses for poorer individuals. The latter provides the economic rationale for defining a given amount of healthcare payments as “catastrophic” and indicative of inadequate financial protection. In this context, the utility losses associated with the
cost of necessary health treatments, relative to an individual’s budget constraint, represent an essential aspect that a more comprehensive theoretical framework for the analysis of health coverage should seek to capture.

An individual utility model

In this section, in contrast to the studies discussed above, we examine the concept of coverage from a conventional micro-economic perspective. We use a simple individual utility function (based on wealth and health) and examine the utility losses that arise as the result of a health shock. We take as a starting point universal free access to all health treatments.\(^1\) We then examine the changes in utility associated with the introduction of a user charge for a given treatment.\(^2\) Such changes in utility for any individual can take one of two forms: a wealth-related loss caused by the need to pay for treatment, or a health-related loss caused by the denial of access to treatment that would otherwise have been secured without the user charge. We make numerous simplifying assumptions in this analysis, but believe it offers a useful framework for conceptualizing and measuring health coverage.

We assume there is a set of \(n\) healthcare problems, and for each problem there is a treatment \(i\) available at a known constant cost \(x_i\). The treatments are assumed to be efficient, in the sense that none is dominated by any other for the specified condition (these are the most cost-effective interventions for each condition). We further assume that the treatments are additively separable, i.e. there are no interactions between diseases or treatments; that the need for treatment \(i\) in wealth group \(y\) occurs with probability \(\pi_i(y)\); and that the health benefits of treatment \(i\) are equal for all groups.

The impact of user charges on utilization can be illustrated by means of a model of individual utility. Suppose all individuals have identical preferences, but differ in health \(h\) and wealth \(y\). Without treatment, a health shock requiring treatment \(i\) reduces health from \(h\) to \(h - b_i\). Without loss of generality, we assume the use of treatment \(i\) restores the individual to health \(h\), so \(b_i\) is the health gain associated with the treatment. The user charge is \(p_i\), the magnitude of which is a policy choice that depends on the level of public subsidy (from pooled public funds) provided towards the treatment. The charge is some proportion (less than or equal to one) of the cost of treatment \(x_i\). We assume a utility function \(u(h,y)\), where \(u_1>0\); \(u_2>0\); \(u_{11}<0\); \(u_{22}<0\).

For those who secure treatment at the (possibly subsidized) price \(p_i\), the wealth-related utility loss, relative to having free access to the treatment, is:

\[
\Delta u_i(h, y) = u(h, y) - u(h, y - p_i) \tag{1}
\]

This gives the loss of utility associated with the imposition of a price \(p_i\). Under the assumptions of a conventional utility function, we can assume that this loss decreases as initial wealth \(y\) increases, and decreases as the price \(p_i\) decreases. Note that at this stage we do not include consideration of the insurance premium or tax payment needed to secure the stated level of coverage. Introduction of user charges is useless without the user charge.

\(^1\) An alternative formulation would be to make the starting point complete absence of coverage, so that access to healthcare can be secured only by payment of the full market price. We would then examine the utility gain from public subsidies of the prices. The analytic substance of the two approaches is identical.

\(^2\) In this paper we adopt a broad definition of user charges referring not only to the direct costs of access to care, but also indirect costs such as those arising from travel to health facilities and time off work.
charges reduces the need for such pre-payments, so in a more complete formulation this reduction represents a benefit arising from direct user charges. The magnitude of such benefit would depend on the distribution of insurance premia and tax payments within the population.\(^3\)

If someone foregoes treatment because of the charge \(p_i\), the equivalent utility loss (relative to free access to the treatment) is:

\[
\Delta_N u(h, y) = u(h, y) - u(h - b_i, y)
\]

This expression reflects a loss of health \(b_i\), the magnitude of which increases as the benefits of treatment \(b_i\) increase. It is not clear, a priori, how this loss varies with income for given levels of \(h\) and \(b_i\). It may, for example, be independent of income. For our model, the only necessary assumption is that \(u_{by} < u_{yy} < 0\). That is, as wealth increases, the marginal utility with respect to health declines at a slower rate than the marginal utility with respect to wealth, an assumption that seems plausible under almost all circumstances.

Assuming all individuals have the same initial health \(h\), and face the same charge \(p_i\), for treatment \(i\), there will then be a critical wealth level \(y_i^*\) for treatment \(i\) for which the two utility levels (with and without treatment at price \(p_i\)) are equal, such that:

\[
\Delta_N u(h, y_i^*) = \Delta_N u(h, y_i^*), \text{ and therefore } u(h - b_i, y_i^*) = u(h, y_i^* - p_i)
\]

For values of \(y\) in excess of \(y_i^*\) the patient will opt to seek treatment, whilst for lower levels of \(y\) the patient will forego treatment. This is illustrated in Figure 2, which shows the utility losses at different levels of wealth relative to free access to treatment. Below wealth \(y_i^*\), the utility loss is lower if treatment is foregone, because the price involved leads to an unacceptable sacrifice of other goods and services (possibly including basic food and shelter for the very poorest). The treatment will be perceived as unaffordable. At higher levels of wealth the utility loss is lower if treatment is purchased.\(^4\)

---

\(^3\) Throughout this paper we assume that a fixed level of public funding is available with which to subsidize certain healthcare treatments. In principle, as for example in Petretto (1999) or Smith (forthcoming), it is possible also to model the optimal aggregate level of coverage, and to accommodate the opportunity cost of public funding. However, that requires strong assumptions about the nature of the tax instruments available to policy makers. This paper instead focuses on the public policy problem of choosing the treatments covered.

\(^4\) In this paper we do not consider uncertainty, either in health outcome or in price. In practice such uncertainty may lead to substantial variations in ex post utility level.
Figure 2: Comparison of utility loss (relative to free access to treatment) if treatment is purchased or foregone at subsidized price $p_i$

Another way of looking at this is to examine, for those with income level $y$, the critical points at which they are indifferent between receiving treatment (with benefit $b$ and payment $p$) and foregoing treatment (declining the benefit, but avoiding the payment). These points are defined by the relationship:

$$u(h-b,y) = u(h, y-p)$$  \hfill (4)

Figure 3 illustrates the locus of pairs $(p,b)$ at which such indifference occurs for two particular wealth groups, described as rich and poor. Any treatment $(p_i, b_i)$ falling in the space to the left of the relevant curve will be purchased, any to the right of it will be declined. The space of acceptable treatments enjoyed by the poor group is of course smaller than the equivalent space for the rich group, reflecting the poor group’s higher marginal utility of wealth. The slope of the curve reflects the diminishing marginal utility of wealth (its gradient is $\partial u / \partial y \div \partial u / \partial h$ at the post-treatment level of wealth). For any particular treatment $i$, the critical level of wealth $y_i^*$ is indicated by the wealth-related curve that passes through the point $(p_i, b_i)$. 
The expected utility loss associated with a price \( p_i \) for treatment \( i \) depends on the probability of occurrence \( \pi_i(y) \) and the wealth distribution of the population \( \gamma(y) \). Assuming a utilitarian model of social welfare, the aggregate societal utility loss associated with user charge \( p_i \) could then be written as:

\[
EUL_i(p_i) = \int_0^{\gamma_i(p_i)} \Delta u_N(h, y) \pi_i(y) \gamma(y) \, dy + \int_{\gamma_i(p_i)}^{\infty} \Delta u_Y(h, y) \pi_i(y) \gamma(y) \, dy
\]

\[
= \int_0^{\gamma_i(p_i)} u(h, y) \pi_i(y) \gamma(y) \, dy - \int_0^{\gamma_i(p_i)} u(h - b_i, y) \pi_i(y) \gamma(y) \, dy - \int_{\gamma_i(p_i)}^{\infty} u(h, y - p_i) \pi_i(y) \gamma(y) \, dy
\]  

The first expression, the level of utility with full access and no user charges, is independent of \((p_i, b_i)\). The second expression gives the expected aggregate utility for those who do not receive treatment at price level \( p_i \), whilst the final expression gives the expected aggregate utility for those who pay for treatment at that price level.

A more general expression for total welfare loss can be assessed by integrating these utility losses into a social welfare function, for example by incorporating a set of weights \( w(y) \) based on wealth levels:
Equation (6) can be thought of as the “expected welfare costs” associated with user charge $p_i$ for treatment $i$. Summed across all treatments, it is in principle this welfare loss that any metrics of health coverage should be seeking to capture. Therefore, in an ideal world, to capture the full effect of a set of user charges (relative to a benchmark of zero price for all treatments) we would probably try to mirror all the utility losses represented in $WL(p_i)$ and then sum across all treatments. In practice this is of course infeasible, but it will be useful to keep this formulation in mind when assessing alternative coverage metrics.

**Existing metrics of health coverage**

Note that the composition of the sub-populations $Y_i$ (those who receive treatment) and $N_i$ (those who decline) will depend on the nature of the price regime. As above, the total annual utilization of treatment $i$ (the size of population group $Y_i$) can be expressed as a function of the payment $p_i$:

$$\theta_i(p_i) = \int_{y_i(p_i)}^{\infty} \pi_i(y) \gamma(y) dy = \sum_{j \in Y_i} \pi_{ij}$$

(7)

The total need for treatment $i$ can be thought of as:

$$n_i = \theta_i(0) = \int_{0}^{\infty} \pi_i(y) \gamma(y) dy = \sum_{j \in Y_i} \pi_{ij} + \sum_{j \in N_i} \pi_{ij}$$

(8)

which is the number of users if there are no charges. Clearly, $\theta_i'(p) \leq 0$. However, we can make no general statements about higher order derivatives.

A commonly used measure of coverage is the ratio of actual to total expected utilization (World Health Organization 2010b), which is an empirical estimation of the ratio

$$\frac{\sum_{j \in Y_i} \pi_{ij}}{\left( \sum_{j \in Y_i} \pi_{ij} + \sum_{j \in N_i} \pi_{ij} \right)}$$

(9)

often referred to as an indicator of unmet medical need. This metric shows the relative size of the groups $Y_i$ and $N_i$ (weighted for disease incidence), and therefore makes reference to the number of

---

In principle, in order to promote equity concerns, it is possible for user fees to be varied across wealth levels according to some function $p_i(y)$. This leads to a more complex formulation of the social welfare function that is not germane to the argument in this paper.
people in need yet foregoing treatment, and the associated loss of health benefits. However, it does not quantify the consequent health-related utility losses, and makes no reference to the wealth-related utility losses associated with user charges.

For treatment $i$, the observed proportion of costs borne by patients out-of-pocket, $OOP_i$, will be as follows:

$$OOP_i = \sum_{j \in Y_i} \pi_{ij} p_j / \sum_{j \in Y_i} \pi_{ij} x_i$$

(10)

where $p_j$ is the price paid by person $j$. Of course, this equals $p/x_i$ if the price is the same for all patients. This expression is an indirect reflection of the wealth-related utility losses associated with treatment, but makes no reference to the differential utility losses suffered by patients at different wealth levels. Also important, there is no reference to the group $N$ in this formulation.

“Catastrophic” payments are incurred by those who make use of the service, but suffer especially high utility losses. The exact formulation of this concept requires a normative judgement about the thresholds beyond which such financial catastrophe occurs. The spirit (if not the practice) of indices of catastrophic spending would require counting those in group $Y_i$ for whom the utility loss $\Delta u_i(h, y) = u(h, y) - u(h, y - p_i)$ is greater than the chosen norm. It therefore requires a count of a subset of the population accessing the treatment:

$$U \int_{\gamma_i(p_i)} \pi_i(y) \gamma(y) dy$$

(11)

where $U$ is some upper wealth limit above which financial catastrophe is not experienced. Notice that, if defined in terms of a maximum acceptable utility loss, the limit $U$ will be implicitly defined by the pair $(b, p)$.

**Towards a utility-based index of health coverage**

In approximate terms, the above equation (6) could be converted into a manageable index of coverage by assuming utility is separable in health and wealth and linear in health (as before). Ignoring equity weights, and assuming a general vector of treatment prices $p_i$, the expected utility loss for treatment $i$ relative to free access is:

$$WL_i(p_i) = \sum_{j \in N_i} \pi_{ij} b_i + \sum_{j \in Y_i} \pi_{ij} [v(y) - v(y - p_j)]$$

(12)

The first expression gives the expected health-related loss suffered by those who would decline treatment $i$ due to user charges. Their loss is simply denoted by the foregone health benefits $b$. The

---

6 Throughout the paper we refer to losses at the individual level. However, it is often more natural to consider losses at the household level, and much of the empirical research uses the household as the unit of analysis.
second expression is the expected wealth-related loss suffered by those who receive treatment and pay \( p_{ij} \). Thus, an approximate operational index of coverage for treatment \( i \) could be constructed by adding up: (a) the number of people \( N_i \) estimated to forego treatment due to price barriers, weighted by the expected loss of health; and (b) the number of people \( Y_i \) who secure access to treatment by paying a charge, weighted by the expected wealth-related utility loss.

Note that the latter part of the index must weight people according to their wealth levels, as the utility loss \( v(y) - v(y-p_i) \) is a declining function of wealth, because of the diminishing marginal utility of wealth. By construction, these losses for group \( Y_i \) are less than the health-related loss \( b_i \). At the critical wealth level \( y_{i^*} \), the utility loss associated with not receiving treatment equals the utility loss associated with paying for treatment, and this utility loss then declines as wealth increases. Thus, just above the critical wealth level, individuals might suffer a utility loss not very different to the health-related loss \( b_i \), suffered below the critical wealth level. The utility loss then diminishes as wealth increases further.

Part (a) requires a count of unmet need. Some studies have attempted to measure unmet need using survey data, based on the amount of healthcare that an individual, on average, would have received had they been treated as other individuals with the same age, gender and morbidity characteristics were treated (Wagstaff and van Doorslaer 2000). Such methodology can serve as a starting point to provide an estimate (albeit probably crude) of the number of people foregoing access to necessary care. A rough estimate of the health benefits lost can be obtained, in terms of quality-adjusted life years (QALYs), from the literature on the value of health gains or related cost-effectiveness analysis.

Part (b) requires a utility-based estimate of the financial hardship caused by securing access to necessary care, with those on lowest incomes weighted most highly. Amongst existing metrics, simply counting the number paying for care or the payments made (as implicit in measures of out-of-pocket spending) is unlikely to be satisfactory. Measures of the numbers suffering catastrophic (or impoverishing) spending caused by the treatment, and in particular “gap” indicators that measure by how much a given catastrophic threshold is exceeded on average in each income group (Wagstaff and van Doorslaer 2003), come closer to offering estimates of wealth-related utility losses, because they reflect the heavier burden borne by poorer people close to the critical wealth level. These numbers would be weighted by some proportion \( \alpha (<1) \) of the QALY loss they would have suffered if they had not received treatment. The value of \( \alpha \) chosen would depend on how tightly “catastrophe” was defined.

Up to now we have considered a single treatment in isolation. However, any comprehensive metric of health coverage should in principle consider all relevant treatments and seek to aggregate their coverage according to some concept of importance to the population under scrutiny. Note that the utility loss described above is defined in terms of QALY loss or its wealth-related equivalent. It therefore automatically reflects the different effectiveness of different treatments, as expressed in potential health gains. Summation of equation (6) across a chosen bundle of treatments can thus serve as a basis for estimating total expected welfare losses associated with financial barriers to that bundle. However, such summation ignores variations in costs of different treatments, and the associated concern of whether the treatments under consideration are the appropriate priorities for a health system, given the treatment costs \( x \) and the limited funding available. Therefore, a signal of the contribution of treatment \( i \) to welfare losses relative to other treatments, given price vector \( p \), would be the expression:
As before, the top line gives the aggregate welfare loss associated with all individuals in need of the treatment, reflected in foregone treatment or imposition of charges. The bottom line weights those losses in relation to the costs of providing treatment. In effect, the welfare loss associated with each person in need of treatment is valued at the treatment's benefit/cost ratio if treatment is foregone, or the equivalent ratio for wealth-related loss if treatment is accepted.

This index can simply be summed across treatments as an estimate of the total utility losses caused by a chosen price regime \( p \), as follows:

\[
TWL(p) = \sum_i \left( \sum_{j \in N_i} \pi_j b_j + \sum_{j \in Y_i} \pi_j \left[ v(y) - v(y - p_j) \right] \right) / x_i
\]

(14)

This gives the magnitude of unmet need across the health basket (weighted for health benefits lost) and the magnitude of finance-related utility loss. The latter may be easier to estimate across an entire health basket rather than for individual treatments. As discussed above, measures of estimated QALY losses due to a lack of access can act as a basis for the health-related utility losses, whereas measures of catastrophic spending could act as (highly imperfect) proxies for the wealth-related utility losses. The elements of this index are weighted according to the inverse of costs of treatment \( x_i \), in order to reflect the relative opportunity cost of removing financial barriers to access.

Finally, it is worth noting that the utility-based index should, in principle, refer specifically to the set of “needed” services. This is not straightforward since the definition of what a needed service means is open to debate and may vary across societies (Culyer and Wagstaff 1993). Moreover, even with no formal coverage, some citizens may secure access to treatment by paying the full market price. A basic healthcare package might be specified by ranking treatments according to their cost-effectiveness ratios, and identifying which treatments should be fully or partially subsidized by pooled resources. In some cases, free access might be limited to pool members or to those who qualify for fee exemptions. In the Appendix, we have proposed an optimization model that seeks to indicate which services should be covered, and to what extent, given the fixed magnitude of the public service budget. The results suggest that: (i) zero user charges should be attached to treatments of high cost-effectiveness, where the price elasticity of demand is such that the imposition of charges would deter many patients or lead to substantial impoverishment; (ii) full charges (or equivalently exclusion from the chosen package of coverage) should be applied to treatments where the public cost of any subsidy would outweigh any benefits in health (for patients now able to secure access) or reduction in impoverishment; and (iii) intermediate copayments are optimal when the incremental benefits of increased charges (reduced expenditure from public funds) exactly balance their costs (increased impoverishment and – at the margin – deterrence of some utilization, with consequent reduced health).
Discussion

There is growing interest in developing consistent indices of health coverage in order to compare different health systems, and to track progress over time. As described in the introduction, the concept of coverage is now usually defined so as to encompass three dimensions:

- The proportion of the population for whom access to healthcare is available from pooled health funds;
- The proportion of needed services covered by pooled funds; and
- The proportion of costs paid from pooled funds.

Various existing indicators offer insights into individual aspects of health coverage. However, our paper suggests that these metrics are either incomplete or misleading for the purposes of providing a broader picture of coverage entailing both access to needed care and financial risk protection. Such existing metrics can, nevertheless, be integrated into a coherent conceptual framework. We propose a formulation based on microeconomic principles that could serve as the basis for an operational index of coverage, reflecting concerns with both unmet treatment need and financial hardship due to healthcare payments.

The framework developed in the paper gives valuable insights into the main areas where existing metrics of coverage could be adapted to better address the fundamental policy concerns of access to necessary care and financial protection. The percentage of the population in need that actually receives treatment when payment is made is an indication of the proportion covered for such treatment, and we indicate that it can reflect the health-related utility losses associated with foregone care by incorporating estimates of unmet need from survey data, and corresponding QALY losses from the related literature. As for the financial protection dimension, metrics that indicate the extent of out-of-pocket payments, and catastrophic or impoverishing payments, normally relate to aggregate payments for health services (some of which may not be considered priorities) rather than charges for specific treatments. Also, such indicators rarely weight for the greater utility losses suffered by poorer people. We argue that measures of catastrophic losses that reflect the different valuations of the degree of loss suffered represent a more promising alternative. Whilst the empirical literature on catastrophic payments is yielding useful insights into wealth-related utility losses, the discussion here suggests that the normative assumptions underlying such metrics could be reconsidered.

As well as informing descriptive metrics of coverage, the model we set out above can act as a normative basis for deciding which treatments should be covered by a limited pool of public finance, and the magnitude of any associated user charges. Equity weights can be readily applied to this model.

Of course the models set out in this paper are conceptual, so there are a number of simplifications involved in order to make the analysis as general as possible, and these point to areas where further research to extend the suggested theoretical framework would be beneficial.

The concept of “need” is not explicitly defined, but is implicitly incorporated as “capacity to benefit” (which could arguably vary across individuals according to factors such as age), possibly weighted by equity. Future work might introduce heterogeneity both in initial health status $h$ and in capacity to benefit $b$. We have also assumed that clinical need is readily observable, and that only those with the presumed capacity to benefit will receive care. In practice, reduction of user charges may lead to moral
hazard or supplier induced demand unless adequate gatekeeping and provider payment incentives are in place. Future modeling might accommodate such issues.

The model makes no explicit reference to uncertainties concerning price and the quality of treatment received. Where quality variations exist, they can be modeled as an adverse impact on measured coverage, and their influence on coverage decisions can also be assessed.

Finally, our current formulation considers the decision to seek care as determined essentially by the effect of user charges: it does not incorporate the influence of non-price determinants of treatment uptake. Some of these are readily accommodated by extending the definition of user charges to embrace implicit costs to patients, such as travel costs and loss of earnings. Other determinants, such as cultural barriers, are less readily conceptualized within an economic framework. Also, we have assumed that the alternative to publicly financed coverage is private spot payment for treatment. If a voluntary health insurance market exists, richer citizens may be able to protect themselves from some of the finance-related disutility associated with health shocks. This possibility is not a widespread phenomenon in low income and many middle income settings, but can be readily modelled if necessary (Smith, forthcoming).

Existing measures of health coverage tend to focus on de jure rather than de facto access to services, or simple counts of the number of people incurring particularly high healthcare expenses. Such metrics have usually been proposed and applied without a firm theoretical micro-economic foundation. We suggest a utility-based framework that could serve as a basis for the development of a meaningful summary measure of de facto coverage. In principle, our theoretical index formulation (and its extension to include non-price barriers) allows the utility loss from deficient coverage to be decomposed into various factors that may be hindering access to necessary care. Ideally, an operational index of health coverage should share this feature in order to maximize its relevance for policy guidance. Although it may eventually prove infeasible to measure explicitly all the relevant elements of coverage in a single index, it seems crucial to start with an integrated theoretical framework that offers a coherent basis for discussing the influence of all relevant coverage dimensions.

References


7 Private prepaid plans represent on average just 3% of total health expenditures in low-income and middle-income countries (data for year 2008 from the Global Health Observatory: http://www.who.int/gho/database/en).


Figure 1: Three dimensions to consider when expanding health coverage (source: World Health Organization 2010a)
A model of optimal coverage

Given the framework set out in the main text, a fundamental question that arises is: what should be the set of treatments that comprise the package of “needed health services” to which citizens should be offered access? In this section, we set out a model that determines the socially optimal set of user charges for a fixed government budget $X^*$, extending the framework suggested by Smith (2005). We assume that the utility function is linear in health (as is implicit in traditional cost-effectiveness analysis) and is separable in health and wealth, such that $u(h,y) = h + v(y)$. The fundamental decision variables for the policy maker are the copayment levels $0 \leq c_i \leq 1$ for each treatment $i$, where $p_i = c_i \cdot x_i$.

There is indifference between treatment and no treatment when:

$$ b_i + v(y - c_i x_i) = v(y) \quad (A.1) $$

At critical wealth $\psi(c_i)$ for treatment with copayment $c_i$, costs $x_i$ and benefits $b_i$:

$$ b_i + v(\psi(c_i) - c_i x_i) = v(\psi(c_i)) \quad (A.2) $$

Demand for treatment is:

$$ \left\{ \begin{array}{l}
\theta_i(c_i) = \int_{\psi_i(c_i)}^{\infty} \pi_i(y)\gamma(y)dy \\
\theta_i(c_i) = -\pi_i(\psi(c_i))\gamma(\psi(c_i))\psi'(c_i) 
\end{array} \right. \quad (A.3) $$

The expected utility loss associated with a set of copayments $c_i$ is then:

$$ UL = \sum_i \left( \int_0^{\psi_i(c_i)} b_i \pi_i(y)\gamma(y)dy + \int_{\psi_i(c_i)}^{\infty} [v(y) - v(y - c_i x_i)]\pi_i(y)\gamma(y)dy \right) \quad (A.4) $$

And the global coverage expenditure constraint reflecting the payments required from public funds is:

$$ \sum_i \int_{\psi_i(c_i)}^{\infty} (1 - c_i) x_i \pi_i(y)\gamma(y)dy \leq X^* \quad (A.5) $$

Then the associated Lagrangean is:
\[ L = \sum_{i} \left( \int_{0}^{\psi(c_i)} b_i \pi_i(y)(y)dy + \int_{\psi(c_i)}^{\infty} [v(y) - v(y - c_i x_i)] \pi_i(y)dy \right) \]
\[ + \lambda \left( \sum_{i} \left[ \int_{\psi(c_i)}^{\infty} (1 - c_i)x_i \pi_i(y)\gamma(y)dy \right] - X^* \right) - \sum_{i} \mu_i (c_i - 1) \]  

(A.6)

Differentiation with respect to copayment \( c_i \) yields:

\[ \frac{\partial L}{\partial c_i} = b_i \pi_i(\psi(c_i))\gamma(\psi(c_i))\psi'(c_i) + \{v[\psi(c_i) - c_i x_i] - v(\psi(c_i))]\pi_i(\psi(c_i))\gamma(\psi(c_i))\psi'(c_i) \]
\[ + x_i \int_{\psi(c_i)}^{\infty} v'(y - c_i x_i) \pi_i(y)\gamma(y)dy - \lambda x_i \int_{\psi(c_i)}^{\infty} \pi_i(y)\gamma(y)dy \]
\[ - \lambda x_i (1 - c_i) \pi_i(\psi(c_i))\gamma(\psi(c_i))\psi'(c_i) - \mu_i \]  

(A.7)

The terms in this expression reflect the consequences of a marginal increase in copayment, in turn as follows:

(a) The loss of health benefits to marginal citizens foregoing treatment;
(b) The gain in wealth-related utility to marginal citizens foregoing treatment;
(c) The loss of wealth-related utility amongst all users of the treatment;
(d) The social gain of reduced public expenditure on remaining users; and
(e) The reduction in social expenditure on the marginal citizens foregoing treatment.

As in Smith (2005), this formulation leads to three classes of solution for the values of each \( c_i \):

treatments for which the copayment is zero, and access can be secured free of charge; treatments for which there is no public subsidy (copayment is unity), and access can be secured only by payment of the full market price \( x_i \); and intermediate solutions for which some (but not full) public subsidy is optimal. We consider the nature of each solution in turn and examine only first order conditions. The functions \( \theta_i(c) \) are not in general concave, so these results only describe necessary characteristics of an optimal solution.

When price is zero, \( c_i^* = 0 \) and \( \mu_i = 0 \) and the optimization simplifies to:

\[ \frac{\partial L}{\partial c_i} = (\lambda x_i - b_i) \theta'(0) - \lambda x_i \theta_i(0) + x_i \int_{0}^{\infty} v'(y) \pi_i(y)\gamma(y)dy \geq 0 \]  

(A.8)

which can be rearranged to yield:

\[ b_i \theta'(0) \leq x_i \left[ \lambda \theta'_i(0) - \lambda \theta_i(0) + \int_{0}^{\infty} v'(y) \pi_i(y)\gamma(y)dy \right] \]  

(A.9)
This expression shows that for a zero optimal copayment the loss of health benefits caused by a marginal increase in price (the left hand side) exceeds the net financial consequences of the increase, reflected in: (a) the direct impact on public funds of reducing the subsidy; (b) the reduction in demand caused by the price increase; and (c) the impact on personal utility caused by increased payment. In short, the benefits of any reduction in public expenditure secured by a copayment are outweighed by the diminished utility levels expressed in terms of reduced health and increased private payment. Price elasticity of demand is therefore the key consideration.

When subsidy is zero, \( c_i^* = 1 \) and \( \mu_i \geq 0 \) and the optimization simplifies to:

\[
b_i \theta_i'(1) \geq \{v[\psi(c_i) - c_i x_i] - v(\psi(c_i))\} \theta_i'(1) + x_i \int_{\psi_i(c_i)}^{\infty} v'(y - c_i x_i) \pi_i(y) \gamma(y) dy - \lambda x_i \theta_i(1)
\]

(A.10)

\[
\int_{\psi_i(c_i)}^{\infty} v'(y - c_i x_i) \pi_i(y) \gamma(y) dy \leq \lambda x_i \int_{\psi_i(c_i)}^{\infty} \pi_i(y) \gamma(y) dy
\]

(A.11)

This indicates that the gains in health derived from a reduction in copayment (the left hand side) are outweighed by the individual and societal costs of the reduction.

When an intermediate solution is optimal, \( 1 > c_i^* > 0 \) and \( \mu_i = 0 \). Then:

\[
x_i \int_{\psi_i(c_i)}^{\infty} v'(y - c_i x_i) \pi_i(y) \gamma(y) dy - \lambda x_i \int_{\psi_i(c_i)}^{\infty} \pi_i(y) \gamma(y) dy
\]

\[
- \lambda x_i (1 - c_i) \pi_i(\psi(c_i)) \gamma(\psi(c_i)) \psi'(c_i) = 0
\]

(A.12)

The expression reflects, in turn, the marginal impact of an increase in copayment on: (a) the loss of wealth-related utility amongst treatment users, due to the increased fees; (b) the associated reduction in public expenditure on all remaining users; and (c) the reduced public expenditure associated with patients deterred from treatment at the margin. In other words, loss of wealth-related utility arising from a copayment increase must equal the social value of the consequent reduced public spending. It can be rewritten as:

\[
\int_{\psi_i(c_i)}^{\infty} v'(y - c_i x_i) \pi_i(y) \gamma(y) dy = \lambda \theta_i(c_i) - \lambda (1 - c_i) \theta_i'(c_i).
\]

(A.13)