WHO/ARA/CC/97.3
Original: English
Distr.: General

CURRENT CONCERNS

ARA Paper number 13

METHODS FOR EVALUATING EFFECTS OF HEALTH REFORMS

Barbara McPake
Health Economics and Financing Programme
Health Policy Unit
London School of Hygiene and Tropical Medicine

Joseph Kutzin
Division of Analysis, Research and Assessment
World Health Organization

Division of Analysis, Research and Assessment
World Health Organization

Within the overall area of Health Systems Development, the Division of Analysis, Research and Assessment is concerned with analytic and normative functions related to the development and reform of health systems in order to support technical cooperation and capacity building in countries. It identifies priority issues in health policy and systems development, supports action to address these issues, and develops, disseminates and supports the application of methods and tools needed in the analysis of policy options and operation of health systems.

The Current Concerns Series is intended to give a wider audience to contemporary issues in the achievement of Health for All as reflected through the work of ARA. These working papers are not official policy statements by WHO, but are summaries of recent experiences based on ARA work in countries. The issues selected are considered to be of immediate concern to countries, and for which there is strong demand for information. Comments and suggestions are welcome and should be addressed to :

The Director
Division of Analysis, Research and Assessment
World Health Organization
1211 Geneva 27
Switzerland

WHO/ARA/CC/97.3
Original: English
Distr: General

CURRENT CONCERNS

ARA Paper number 13

METHODS FOR EVALUATING EFFECTS OF HEALTH REFORMS

by

Barbara McPake
Health Economics and Financing Programme
Health Policy Unit
London School of Hygiene and Tropical Medicine

and

Joseph Kutzin
Division of Analysis, Research and Assessment
World Health Organization

World Health Organization

TABLE OF CONTENTS

Acknowledgments

We gratefully acknowledge the helpful comments and contributions received from Orvill Adams, Pascal Brudon-Jakobowicz, Guy Carrin, Andrew Creese, Katja Janovsky, John Martin, Jonathan Quick, Eleuther Tarimo, and Phyllida Travis.

The Health Economics and Financing Programme of the London School of Hygiene and Tropical Medicine is supported by the Overseas Development Administration (now the Department for International Development) of the United Kingdom.

1. Introduction

This paper aims to identify simple methods to evaluate effects of health sector reforms and to discuss the relative usefulness of various methods in different situations. We consider methods that attempt to measure directly the effects of changes on health system objectives. These methods are intended to contribute to a broader process of evaluation that includes analysis of the context in which reforms are introduced and the process by which they are designed and implemented, in addition to the measurement of the ultimate effects of reform and an assessment of the policy implications of these findings.

1.1 The context of health sector reform

Health systems in developing and transitional countries have been subjected to a variety of pressures and have undergone many changes in recent years. Although often overlooked, evaluation has been needed to assess changes in the light of the objectives of health systems. In this decade however, a substantial number of countries have considered and in some cases started to implement fundamental reorientation of their health systems which goes far beyond the piecemeal and single policy change which characterised earlier decades. For example, the overall structure and organisation of provision of health services are being reexamined, and the ideological orientations of systems are being called into question. The term 'health sector reform' has been coined to reflect the new context.

Cassels (1995) suggests a six-part categorisation of the components of reform programmes: *improving the performance of the civil service* which reduces the constraints within which the health sector functions; *decentralisation* which increases the autonomy of managers at lower levels of the system; *improving the functioning of national ministries of health* through improved structures and managerial and planning procedures; *broadening health financing options* by introducing new mechanisms such as user fees; *introducing managed competition* most usually through contracting within the public sector or between the public and private sectors; and *working with the private sector* by developing and supporting a specific and complementary role for private providers.

Experience in most of these areas is limited, both in terms of the number of countries which have adopted specific measures and in the time period since implementation in those countries. In some policy areas, most notably that of introducing managed competition, measures have been inspired by those being adopted in industrialised countries, but generalizing from industrialised country experience to other country contexts should only be done with great caution (Collins, Green and Hunter 1994). Experience with introducing user fees has probably been most widespread, but even here, much remains unknown about their implications and the circumstances in which different measures have the best chances of success (McPake 1996). Evaluation of reform measures is needed so that governments can determine the effects of their policies and whether, when, and how certain measures may need to be adjusted. Globally, it is essential to document this accumulating experience so that countries considering particular reforms to their health systems can learn from both the mistakes and achievements of others.

1.2 Purpose and scope of the paper

Our 'target audience' is the large number of government planners and non-government consultants who advise health policy makers and who prepare policy briefs on the basis of their analyses of current situations. While the indicators mentioned in the annex of this paper may be of use to those looking to develop a system for routine monitoring of health system performance, this is not our purpose. Instead, we focus on relatively low cost and non complex evaluation methods that can be applied with limited specialist input, yet that still have the capacity to guide national and sub-national evaluative studies. This focus is driven by a belief that the most relevant evaluations, in terms of their potential impact on management and policy decisions, are those that are (1) implemented by locally-based evaluators on a relatively short time scale and (2) repeatable on a regular basis, or even capable of being integrated into routine monitoring systems. To meet these criteria, evaluations must rely, for the most part, on data that are generated through a country's routine information system. Finally, the emphasis on relatively simple and low cost methods also derives from a belief that in many countries, much of the data needed to assess performance in a meaningful way exists, but these data are not organized for the purpose of addressing specific policy and managerial concerns. This paper suggests ways that such data can be organized, analysed and presented for these purposes.

Health reforms have single or multiple objectives which may be of a political, economic, or public health nature. While we recognize that political objectives are important, our aim is to help analysts to evaluate reforms relative to broad economic and public health objectives. In particular, the objectives we highlight are allocative and technical efficiency (including quality and client satisfaction), equity in access to health services, equity in the finance of health services, and financial sustainability. Reforms could have impacts on most of these objectives, either because the objective is the main purpose of the reform (e.g. financial sustainability is usually the primary purpose of the introduction of user fees), or because there may be side-effects of the reform, which could be positive or negative (e.g. an increased reliance on private sector provision and financing may threaten the access to care of disadvantaged groups). Both types of impact should be evaluated to determine if the extent to which objectives have been achieved justifies the extent to which unwanted side effects have been caused. A preliminary assessment of which kinds of impact are likely to be most important for each reform measure has to be made in order that appropriate methods and a manageable number of indicators are selected to monitor the impact of any particular policy. There is a need, therefore, for the analyst to think carefully about the likely effects of a reform (and to generate testable hypotheses, if appropriate) and to consider the possibility of factors other than the reform being analysed that might also affect the indicators to be measured. This process should also lead the analyst to avoid measuring things that are not relevant to the reform being evaluated. For example, a decentralisation policy may specifically aim to reduce inequities in finance where certain parts of the country have not been receiving their fair share of resources. In this case, the extent to which the desired transition in resource allocation patterns is achieved needs to be evaluated. Where it is intended that existing allocational patterns are maintained but resources simply managed at a lower level of the system, this aspect of an evaluation is unlikely to be warranted.

While the categories of health sector reform referred to above provide a useful framework for understanding the range of possible policy reforms, the methods we describe in this paper are appropriate for the analysis of specific measures or elements of these broad categories. Thus, the

paper is intended to help analysts evaluate the effects of relatively specific or 'narrow' reforms (e.g. a change in user fee policy), or specific features of a programme of reforms, rather than the entirety of a multifaceted or prolonged reform process.

The interpretation of changes in performance indicators is, in many ways, more of an art than a science. In other words, just because an indicator changes does not mean that this change was *caused* by a specific reform, not does it imply, as indicated above, that performance has improved or deteriorated. Thus, evaluating the effects of reforms involves more than tracking changes in one or several indicators; it requires judgment. Nevertheless, judgments as to the cause of observed changes in indicators can be informed by the use of structured methodologies (i.e. a scientific approach). In this paper, we present analytical approaches and methods that aim to enable the evaluator to make a more confident judgment of the extent to which trends in indicators and differences in indicators emerging from comparisons can be *attributed* to the specific reform being analysed. The extensive annex to the paper presents a number of indicators (and categories of indicators) relevant to the health sector objectives described above. The annex also provides guidance on the interpretation of these indicators and suggestions of possible sources of data.

2. Linking Effect with Cause: Basic Approaches to Evaluation

Evaluating the effects of a reform involves describing a policy change, describing (and hopefully measuring) changes in health system performance, and assessing the extent to which the changes observed can be attributed to the reform that was implemented. As noted by Janovsky and Cassels (1996), this is a difficult and challenging task because reforms are not implemented in a laboratory. Policy change is often part of a continuum rather than a discrete event, and sectoral objectives are affected by a wide range of policy and nonpolicy contextual factors that do not stop having their effects simply because a new policy is being implemented. In general, the more complex the policy or the policy environment, the more difficult it is to determine causal links between reforms and health objectives. Therefore, the approaches that we present are not tools of 'hard' science. Instead, they are ways of structuring an analysis to reach *plausible* conclusions about cause and effect, rather than methods that will lead to a *proof* of causality.

As noted in the introduction, reforms should be assessed in terms of their implications for health sector objectives, such as efficiency, equity, sustainability, etc. In the annex, we discuss the extent to which specific indicators reflect accurately the intended objective of policy. Just because an indicator changes, however, does not imply that the change was caused by a change in policy. In this section of the paper, we present methodological approaches that can be used to help evaluate whether, and to what extent, changes in indicators can be associated with changes in policy. We illustrate these approaches with examples from evaluative studies of the effects of reforms in a number of countries. These examples also suggest ideas for the (graphic and tabular) presentation of data analysis to policy makers.

2.1 Descriptive analysis

The identification and measurement of performance indicators are only part of the process of evaluation. As suggested in recent 'frameworks' (Janovsky 1995; Kutzin 1995), the first steps in the process of evaluating reforms are to provide clear and detailed descriptions of (1) key contextual factors driving reform, (2) the reform itself and its objectives, and (3) the process by which the reform was (is being) implemented. Descriptions of the features of policy mechanisms and their implementation can be considered 'descriptive indicators'. If policy reform has involved the introduction of the mechanism where it was previously not used, there is no question that these indicators reflect something associated with the policy reform. For example, if user charges have been introduced, the percentage exempted cannot but be associated with the policy change (although what this implies about access needs further thought--see annex section A.3). This may seem a frivolous point, yet such indicators already enable some analysis of the effects of policy by describing salient features of the policy and its implementation. Many evaluative reports of reform go no further than to describe analytically such features and their effects but still are able to identify many of the policy's strengths and weaknesses and even suggest measures to improve on the policy's performance. A study of decentralization in Tanzania (Gilson, Kilima and Tanner 1994) is an example of an analytic descriptive evaluation that yields significant policy recommendations, and a study of hospital autonomy in Kenya (Collins, Njeru and Meme 1996) also relies very heavily on descriptive analysis as a basis for the authors' conclusions.

Yang (1991) is able to identify problems of cost inflation, inequity, and inefficiency of administration mainly by thinking through the implications of several 'descriptive indicators' (although some other approaches are also employed) in an evaluation of the national health insurance system in Korea. Table 1 shows the indicators used to assess each of these issues.

Table 1: Use of descriptive indicators to evaluate a national health insurance system

Problem (related to objective)	*Descriptive indicator (feature linked to the problems)*
Cost inflation (allocative and technical efficiency)	Reimbursement mechanism (retrospective reimbursement on cost-plus basis)
Inequity (in access and finance)	High co-insurance rate Unofficial 'two-class' health care system Identification of inequitable risk pooling
Inefficiency of administration (allocative and technical)	Unaccountable management duplicated in each society Proportion of administrative costs to total revenue

Moens (1990) also used a number of descriptive indicators to assess equity of access and financial sustainability in an evaluation of a local prepaid health plan system in a Zaïrian health zone (Table 2).

Table 2: Use of descriptive indicators to evaluate a local prepaid health plan

Problem (objective)	*Descriptive indicator (of performance relative to objectives)*
Equity of access	Membership rates and distribution
Financial sustainability	% cost recovery

While descriptive indicators can *sometimes* be used directly for the purpose of analyzing reforms, a clear understanding and description of a reform and how it was (or is being) implemented is *always* needed before one can reach reasonable conclusions about whether any change in a performance indicator is caused by the reform. Several techniques are described below that can be used to help determine the effects of reforms, but they will be of little value unless the evaluator understands and is able to present clearly the content of a reform and the process by which it was implemented. It is of great help to orient this descriptive analysis around a clear conceptual framework that helps to identify critical policy issues and questions (for examples of such frameworks, see Janovsky (1995), Maxwell (1996), and Kutzin (1995)). Without this, the 'descriptive analysis' can easily become a long, unfocused narrative.

2.2 *Methods for making a more convincing evaluation*

In addition to using descriptive analysis as an aid to drawing conclusions about the effects of a policy change, there are two approaches to associate changes in indicators with changes in policy:

- "Longitudinal" approaches compare the same units of observation over a period in which policy changes. For example, the same health facilities might be compared before and after the introduction of a user fee. Analysis attempts to assess the extent to which changes in indicators between the pre- and post-policy introduction periods can be ascribed to the policy.

- "Cross-sectional" approaches rely on there being the opportunity to compare different units of observation (for example health facilities, areas, individuals) among which there is a difference in policy, at the same time. For example, health facilities in which user fees have been introduced can be compared to health facilities in which they have not. Analysis attempts to assess the extent to which differences in the indicators between the two groups of health facilities can be ascribed to the policy.

These approaches, which can be used separately or in combination, can be incorporated in the design of evaluation studies as a means of increasing confidence in conclusions about whether a change in indicators was caused by a change in policy.

Clearly, there are many variations within each of these approaches. For example, instead of comparing the absence or presence of fees, a range of observations with different levels of fees might be used. A potentially powerful approach is to combine both types (longitudinal and cross-sectional) of approaches, by comparing trends between observations where the extent of policy implementation differs. This can control for the effects of factors other than the policy change that might affect indicators in all locations at the same time. For example, if there is a malaria outbreak, utilisation levels in all health facilities are likely to increase. Without knowledge of the outbreak, a researcher who looked at trends only in facilities in which a reform had been introduced might conclude that the reform had caused the increase. But if she also included in her analysis facilities in which the reform had not been introduced, the researcher would be able to see that utilisation increased everywhere and, therefore, that the reform was probably not the cause.

Such studies sometimes result from controlled experiments in which a policy is introduced on a selective basis deliberately so that its effects can be measured, or they can result from 'natural experiments' in which there is an external reason for applying the policy in some places only. Longitudinal comparisons can be "prospective" (when it is possible to start measuring indicators before implementation of a policy), or "retrospective" (when routinely collected data relating to past experience are analysed after policy implementation).

Difficulties in establishing that there is a causal link between changes in policy and changes in indicators are common to both cross-sectional and longitudinal approaches. The underlying problem is that most available indicators are not only affected by the policy under consideration

but also by other policies and/or changes in underlying conditions. As a result, analysis has to be "multi-variate": it must consider the full range of variables which might affect the indicator of interest and ensure that the policy variable can be isolated as the causative factor. For example, if utilisation levels are found to be lower in facilities with user fees than without, it must be established that the explanation is not really a difference in the size or dispersement of the catchment population, in disease profile, or in other factors such as perceived quality of services which affect the popularity of the facilities and which also differ between the two groups.

It is not our purpose to describe the full range of techniques available to address this problem. Instead, we describe a few simple and easily understood methodological approaches and give some examples of their use. Nevertheless, in some cases these methods will not enable strong conclusions about the effects of policy to be reached, and further progress can only be made by employing a statistical expert. Irrespective of the sophistication of the techniques used, however, the attribution of causality for observed effects ultimately requires judgment on the part of the evaluator. The purpose of using the various techniques is to enable better-informed judgments to be made.

Longitudinal (trend) analysis. If a policy change is discrete (for example a new programme is introduced on a specific date), a very simple method of attempting to relate longitudinal data to a policy change is to identify the date of introduction and look for sudden discontinuities or reversals in trends which are then highly likely to be explained by the new programme. An example of this approach (analyzing the data in Table 3) is provided by Moens (1990) in the study of a Zaïrian prepayment scheme mentioned above.

Table 3: Patient revenue and operating cost - trend before and after the year (1986) in which a prepaid health plan was introduced

Year	*Patient revenue*	*Operating cost*	*Patient revenue/cost*
1984	668,449	1,853,629	0.36
1985	878,583	2,035,735	0.43
1986	1,918,905	3,141,105	0.61
1987	3,848,136	4,674,026	0.82
1988	8,034,130	9,909,054	0.81

Although the trends suggest revenue and operating costs were already increasing before the introduction of the plan in 1986, a doubling of revenue each year since introduction compares to an increase of only 30% the previous year. This is strong evidence of an association. Nevertheless, there is still a need to ask whether or not other changes took place at the same time as the policy change which were also discrete. For example, the change in question might be part of a package of reforms introduced at the same time, or might be associated with a change in other government measures that might explain sudden changes in other trends. In this case, an

important factor affecting both operating costs and revenues was inflation, which was nearly 100 percent in 1988. Because inflation affects both costs and revenues, the impact of the policy change on financial sustainability can best be examined by analyzing the change in the cost recovery ratio (last column). The changes in this proportion over time suggest that the introduction of the prepayment scheme probably had a significant impact on cost recovery.

Overall trends also put the impact of a policy change in perspective. For example the effect of changing the reimbursement rules applying to caesarian section and normal delivery in Brazil (with the objective of reducing the rate of Caesarean births) was studied by Barros, Vaughan and Victora (1986). Although there was a decrease in the rate of Caesarean sections in 1980 after this change was implemented in the town of Pelotas, the longer trend depicted in Figure 1 reveals it to be a trivial impact relative to the overall pattern of change.

Figure 2. Trends in Caesarean sections - Pelotas, 1975-1984

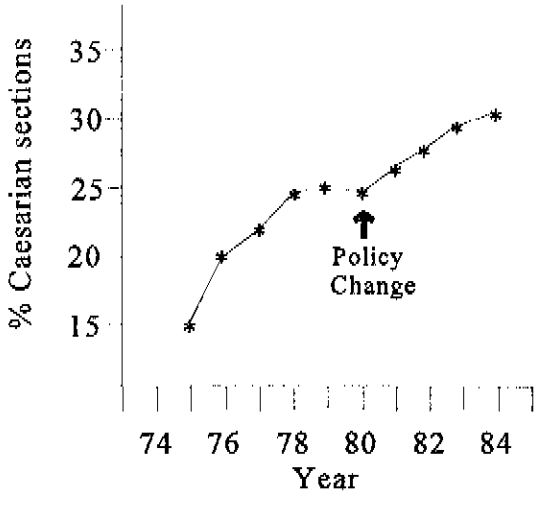


Figure 1 uses a visual technique that is very useful for presenting and understanding the association between policy change and changes in indicators in circumstances of a discrete policy change. This 'mapping' (i.e. explicitly identifying on the graph) of when the policy change took place has proven quite useful in a variety of studies. For example, Waddington and Enyimayew (1990) mapped the timing of the introduction of an increase in user fee levels in Ghana onto graphs depicting quarterly utilisation levels to show the association between utilisation patterns and a price increase. Similarly, Quick and Musau (1994) mapped a series of user fee policy and implementation changes in Kenya onto graphs depicting quarterly revenue collections and utilisation levels. This descriptive technique is a way to show when policy changed and when indicators changed. By using readily available data over several periods before and after the policy change, reasonable conclusions can be drawn that take into consideration the effects of both long term and seasonal trends, without the need to make use of sophisticated statistical techniques. Any observed association does not prove that there is a causal link, but certain patterns (such as the lack of an association) may allow for some possibilities to be ruled out or for conclusions about the effect of the reform to be refined.

A less satisfactory approach to evaluation of the association of a trend with a policy change is exemplified by a study by Yoder (1989) who assessed the utilisation impact of the substantial increase in government user fees in Swaziland (Table 4). Yoder compared the pre- and post-change data without reference to general trends over a longer period. The extent of reduction in utilisation suggests cause and would be unsustainable as a long term trend. Nevertheless, the argument would have been considerably strengthened by more historical information.

Table 4: Monthly average attendance before and after government fees increased to mission fee levels

Sector	*Pre-change attendance (10/83-12/83)*	*Post-change attendance (10/84-12/84)*	*% Change*
Government	817	552	-32.4
Mission	783	862	10.1
Totals	805	665	-17.4

Similarly, Yang, Lin and Lawson (1991) report that in China, following the introduction of payments to staff to work extended hours, the number of monthly surgical operations in one hospital increased from 50 to 80, and following the opening of enterprise-based hospitals to the public, bed utilisation rates increased from 40 to 70%. Again, information about longer term trends would be useful, but the degree of change suggests the observations are unlikely to reflect a long term phenomenon.

It is possible to improve on both these approaches by attempting to identify whether or not there are other policy or environmental changes (i.e. contextual factors) which could explain trends in data other than the financing policy in question. If potential alternative causes are first identified and then ruled out, the case for linking policy and indicator change is strengthened. For example, Yang's (1991) review of the Korean health insurance system, discussed above, reviews a range of possible factors explaining health spending increases (such as general price inflation in Korea and failure to implement adequate controls over technology adoption), before attributing a share of the inflation to the expansion of the insurance system and some of its specific features.

Cross-sectional analysis. Cross-sectional studies resulting from controlled experiments have the advantage that assignment of cases to intervention and control groups can be done randomly, or by a method structured to ensure that differences in results can be explained by the intervention rather than by other predictable factors. It can therefore more safely be assumed that this is the case.[1] In practice, however, controlled experiments are relatively rare, since reforms tend not to be implemented in this manner unless they are designed explicitly as pilot projects.

[1] Evaluations of such experiments also have a disadvantage. Because of the special circumstances under which 'experimental' reforms are implemented, the findings are unlikely to be directly generalizable. See section 2.3 below for more on this issue.

It is much more difficult to reach a firm conclusion on the basis of cross-sectional studies where assignment of observations to intervention and non-intervention groups has not been controlled but rather has been part of the outcome of the policy development process. In these circumstances there will often be a host of factors which differ between the intervention and non-intervention situations and which may have influenced the policy process and thus explain the adoption and non-adoption of the policy. This is not meant as a recommendation as to the desirability of various types of studies, much less ways of implementing reforms. It is simply important to recognize the circumstances under which a reform was implemented in order to make an appropriate interpretation of the information collected in the evaluation study. In cases where differential implementation of reform between different regions or health facilities is an outcome of the way the overall policy is implemented, it is easy to confuse the effects of the reform with the underlying factors that enabled one region or facility to implement the reform first. Interpretation in these situations therefore has to be extremely cautious.

A good example of the problem is a study of the success of trust hospital policy in the UK (Bartlett and Le Grand, 1994). Although trust hospitals exhibited lower unit costs than others, the adoption of trust status was optional, and the authors conclude that hospitals which were already more efficient may have been more attracted to apply for trust status. This aspect of the way the policy was implemented meant that the authors could not attribute the observed differences in indicators (i.e. lower unit costs in trust hospitals) to the policy change.

Combined approaches. Litvack and Bodart (1993) took advantage of a phased implementation of a policy change and selected five facilities, three 'treatment' and two 'control' to evaluate the impact of a user fee accompanied by quality improvement interventions (specifically, a more reliable drug supply) in one province of Cameroon. This approach made possible a 'controlled' method for analyzing a 'natural experiment'. Comparability with the treatment facilities was the principal criterion for selection of the control facilities. This study also had a longitudinal component (baseline information was collected) and did use complex multi-variate techniques to support the conclusions. Nevertheless, a simple comparison of utilisation rates between the 'experimental' and 'control' health centres (the cross-sectional element), before and after (the longitudinal element) the introduction of the user fee/quality change in the experimental health centres makes a convincing case with respect to the impact of the policy change in this region of Cameroon. In the 'control' group (no intervention) utilisation fell while in the 'treatment' group (with the intervention) it increased (Table 5).

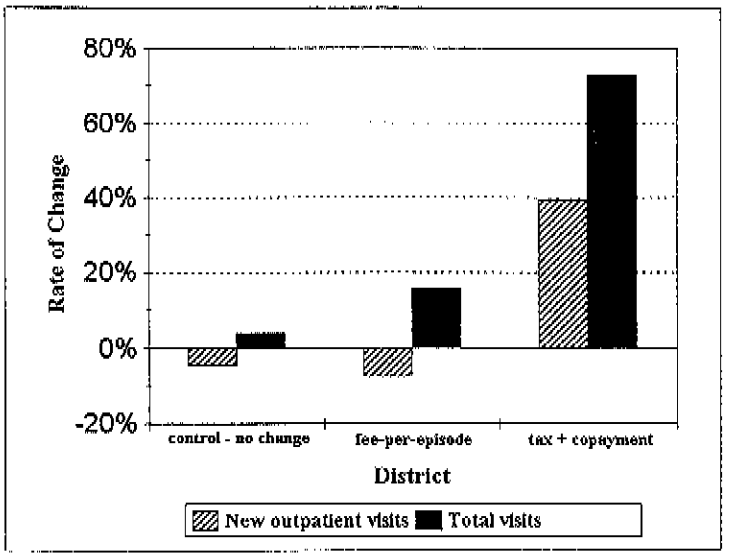
Table 5: Percentage of sick people using a health centre before and after user fee introduction accompanied by quality improvement

	Baseline (%)	*Follow-up (%)*
Control	45	38
Treatment	44	48

A similar methodological approach was used to analyse the effects of alternative cost recovery schemes in three districts in Niger (Diop, Yazbeck and Bitran 1995). The districts had similar economic, social and demographic characteristics (though differing in their ethnic composition), suggesting that it is probably reasonable to attribute any changes in performance to the introduction of policy changes rather than to other factors. In the 'control' district, no change was introduced. In one 'experimental' district, a compulsory health tax on all households was introduced, together with low levels of copayments at health facilities. In the other 'experimental' district, user fees were introduced (fee-per-episode, at higher rates than the copayments in the other experimental district). In both experimental districts, quality and management were enhanced through provision of initial stocks of essential drugs, the introduction of standardized diagnostic and treatment protocols, and the introduction of financial and drug stock management systems.

The indicator that they used to measure the utilisation impact of the alternative fee/quality interventions was *the number of visits to public health facilities*, calculating the percentage change from the year before the intervention to the year of the intervention. As in the Cameroon study, they found that when quality was enhanced, this effect outweighed the effects of a higher user charge. The compulsory health tax and small charge performed better than the full user charge in terms of utilisation. Figure 2 shows how the authors presented their findings.

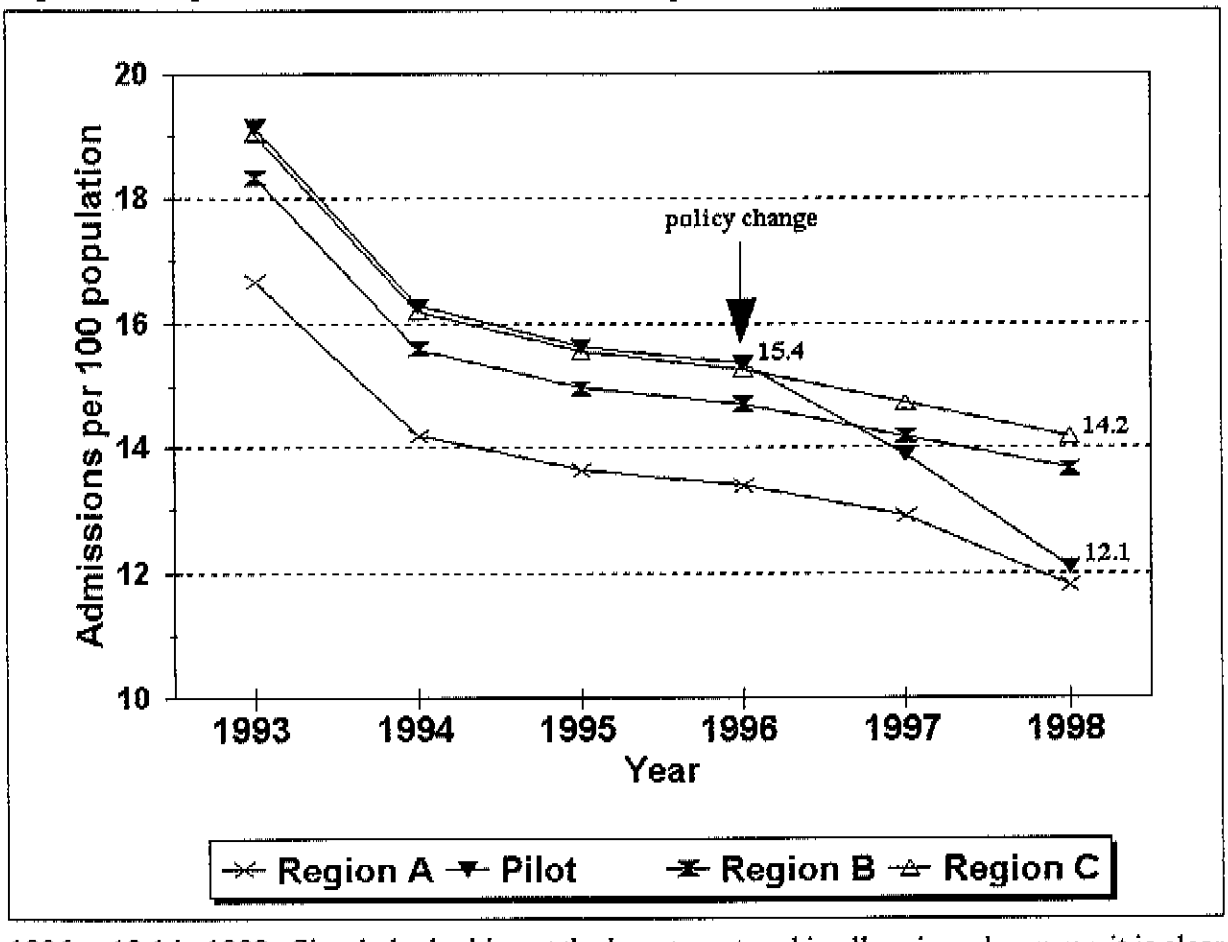
Figure 3. Change in Utilisation in Three Districts in Niger after Different Payment/Quality Interventions



Combining cross-sectional with longitudinal methods covering a longer period of analysis can help to address difficulties in attributing effects to specific policy changes in the context of 'non-random' policy implementation (e.g. explicit pilot projects). This 'combination' involves analysis of data on a given indicator or indicators in both the reforming and non-reforming facilities, districts or regions (i.e. cross-sectional data) for several time periods before and after

the introduction of the reform (i.e. longitudinal or 'trend' data). With such data, one can assess whether the *rate of change* in the reforming facilities/regions differs greatly from that of the non-reformers. For example, Kyrgyzstan has an inefficiently high rate of admission to hospitals as a consequence of multiple factors, and the country has implemented a pilot project in one region to restructure primary health care, retrain providers to manage cases at the first contact level, and reform provider payment to create financial incentives to reduce referrals for inpatient care. One indicator to measure the effects of such changes is the hospital admission rate. However, because the selection of the pilot region was not random, simply comparing its admission rates to other regions after the reform is implemented would not be sufficient to attribute causality. Also, data indicate that, throughout the country, admission rates have been falling for the last several years. Thus, simply comparing admission rates within the pilot region before and after the reforms are implemented would also not be enough to attribute causality. However, combining the cross-sectional with the longitudinal data would allow for a reasonable attribution of the extent of the change that is due to the reforms. Figure 3 uses hypothetical data on admission rates in several regions to illustrate how this might be done. The package of reforms was implemented in 1996 in the pilot region. The admission rate in the pilot region is hypothesized to fall from 15.4 in

Figure 4. Hospital Admission Rates, Selected Regions in Kyrgyzstan (Hypothetical)



1996 to 12.1 in 1998. Simply by looking at the long term trend in all regions, however, it is clear that not all of such a decline would be attributable to the reforms. It is only by seeing that the admission rate is falling faster in the pilot region than in the other regions that we are able to conclude that the reforms as a whole are causing a change in the indicator. Based only on this

indicator, however, we would not be able to attribute this change to one of the specific reforms (i.e. restructuring primary health care, retraining providers, or changes in provider payment) that was implemented. For this, descriptive information on the content and process of implementation of these reforms would be needed.

2.3 Generalizing from the findings: some caveats

Suppose that, after defining a methodology and then collecting and analyzing data, an analyst concludes that a policy change was responsible, at least in part, for a change in an indicator of a health system objective. For example, assume that one concludes, as did the authors of the studies from Cameroon and Niger described above, that a mixture of cost recovery (fees or prepayment) and quality improvement (better drug availability) implemented in one district led to an increase in utilisation. Should this conclusion *automatically* lead one to recommend that the same policy be implemented nationwide? The answer is no. Reaching a conclusion about the effects of the reform being evaluated is a different step, conceptually, from determining the extent to which this conclusion can be *generalized* to other facilities, districts, regions, or countries. This process is also referred to as determining the 'external validity' of the study findings.

In their review of health policy and systems research, Janovsky and Cassels (1996) point to the limitations of generalizing from evaluations of health reform pilot projects or studies of controlled or natural experiments. The same factors that allow for a relatively clear determination of the causes of observed outcomes in these studies (comparison between experimental and control groups, as in the studies from Cameroon and Niger referred to previously) tend to limit the direct applicability of these findings to the health system as a whole. The reason for this difficulty in moving from small-scale pilots to national implementation is the different context that exists with larger scale:

> ...as their scale and scope increases, programmes become more complex; require more coordination; greater commitment of resources; and have wider political implications. The context in which they operate changes and new social, political, economic and organisational factors affect what can be done. Large scale implementation requires facing the structural and system-wide issues from which small-scale projects are effectively protected (Janovsky and Cassels 1996, p.15-16).

This does not mean that small-scale pilots are of no use for informing policy decisions; indeed, they can play a very useful role, especially if they are implemented as part of the national policy development process. In particular, they can be useful for demonstrating what is possible and for identifying at least some of the conditions needed for successful implementation. What they tend not to do, however, is to indicate how these conditions can be put into place as a basis for 'scaling up' or 'rolling out' the reform on a wider (or national) basis. Thus, it is essential to recognize the limitations of this type of study so that national policy recommendations are not made without adequate consideration of the additional requirements of expanded implementation. It is notable that the authors of the Cameroon study (Litvack and Bodart 1993) warned against generalizing their results to areas that do not have similar characteristics to the district in which

they conducted their study. Unfortunately, others have been less careful in generalizing from this work.

One factor that has played a role in the 'success' of many pilot projects in health reform is financial and technical assistance provided by donor agencies. In the Cameroon study, USAID provided management training and an initial stock of drugs. Similarly in the Niger study (Diop, Yazbeck and Bitran 1995), USAID provided initial drug stocks, training in the use of standard diagnosis and treatment protocols, a drug stock and financial management system, and advisors to augment supervisory capacities in the two experimental districts. Obviously, this suggests that more than just the price/quality reform may have been responsible for the effects that were observed. Nevertheless, this does not mean that there is nothing to learn from these studies. Instead, by explicitly recognizing the role of external technical and financial support, the analyst should be able to identify precisely the conditions needed to successfully implement reforms. The authors of the Niger study provide a very clear description of this, leading them to conclude that for Bamako Initiative-like financing schemes to be effective at providing sustainable access to good quality care in more than just a few isolated districts, countries must devote substantial attention to national drug and human resource policies.

Ultimately, as with the evaluation of the effects of a reform, determining the extent to which the findings of an evaluation study are generalizable for broader implementation is a qualitative judgment. This judgment can be enhanced with a clear understanding of the context within which a reform was implemented and the conditions, in addition to the reform itself, that were probably responsible for the observed results. In addition, the analyst must try to identify the additional costs and institutional arrangements needed for broader or national implementation, the likelihood that the necessary changes can be put into place effectively, and the expected time frame for implementing this change.

2.4 Conceptual steps

In this section of the paper, we have attempted to illustrate methods for establishing a causal link between observed changes in performance indicators and reforms that decision makers might wish to see evaluated. The methods presented do not include more complex statistical techniques. Whatever methods are used, it is important to recognize that the attribution of causality will always be a judgment based on probability rather than proof. Still, the methods suggested here can help to increase the evaluator's confidence in his or her conclusions regarding the effects of reform. Based on the methods described in this section, the following series of steps are suggested as an approach to evaluating reforms:

1) **Define what you want to study.** In other words, what is the reform(s) to be evaluated?

2) **Formulate hypotheses/research questions and indicators.** What are the expected effects of the reform(s) on health system objectives? What indicators will be selected to measure these effects?

3) **Identify alternative causes of effects.** Apart from the reform, what else might affect the indicators to be measured? How can the effects of these other factors be accounted for and disaggregated from the effects of the policy change?

4) **Define methods.** Based on steps two and three, and an assessment of available data and the resources available to conduct an evaluation study, define whether the methods will be descriptive only or will incorporate comparisons that are cross-sectional, longitudinal, or a combination of the two.

5) **Describe the policy change and its implementation.** A link between cause and effect cannot be made without a clear description of the content and timing of the implementation of the reform being analysed. This description of the process by which the reform has been (is being) implemented is the essential first part of the analytical work of the evaluation.

6) **Data collection and analysis.** Collect the information on the selected indicators, and analyse it according to the methods defined in Step four. Document the changes (if any) or comparisons in the indicators.

7) **Assess causality.** Based on the description of the implementation process and the methods used, assess the likely causes of the observed changes or differences in performance indicators. Reach a conclusion on the extent to which these changes or differences were caused by the policy change or by other factors. Accept that this is an informed judgment rather than a 'scientific' certainty.

8) **Assess policy implications.** Based on your assessment of the effects of the reform being evaluated, an analysis of the role of contextual factors, the nature of the reform implementation (e.g. pilot experiment or national policy change), and an assessment of the additional changes needed for a wider application of the policy reform, make an assessment of the extent to which your conclusions on the effects of *this* reform are valid for other parts or the rest of the health system. Again, this will be a judgment. Identify the policy, institutional, and management changes needed to 'scale up' from pilot to broader national implementation.

3. Conclusions

In this paper, we have presented broad approaches and specific examples of methods to evaluate health reforms. This is generally the most rigorous and testing form of evaluation. There are indirect types of evaluation which have not been discussed but which in many situations may be all that is possible or most appropriate.

Some types of policy are likely to be more difficult to relate to changes in objectives and might be better evaluated through an indirect approach. It will always be difficult to relate changes in an indicator of a policy objective to a reform process that is multifaceted and evolves in a slow and incremental manner. An example is decentralisation policy which can incorporate many changes that occur over a long period of time. It is unlikely, therefore, to achieve noticeable effects on performance indicators in the short term. Similarly, the effects of measures to strengthen institutional capacity evolve slowly. If such policies are successful, the results will gradually filter through to indicators of health sector objectives. In such cases, there are arguments for measuring change further back in a theoretical chain of causation from the policy change to the objective. For example, Appleby *et al.* (1994) evaluated the development of managed competition in the UK context by gathering data on the details of business plans, the composition of contracting teams, types of contracts and contract negotiations, reform implementation arrangements and issues concerning staffing, skills and expertise demanded by the reforms as perceived by purchasers and providers.

Measuring the views of those involved either in implementing reforms or in receiving services is a common approach in such circumstances (for example Ruwe, Macwan'gi and Atkinson 1996). Even in situations in which quantitative data are quite reliable, such methods may dominate those used to evaluate reform. In a review of seven studies evaluating the impact of 'GP fund holding' in the UK on equity (Whitehead 1994), for example, only two used quantitative indicators.

All such evaluations are less conclusive about the effects of reforms, and, arguably, less useful than those which aim to evaluate changes in health policy objectives. While they are often more appropriate for policies at early stages of implementation, or ones which are only expected to have a very indirect impact on policy objectives, the ultimate test of reform is to show that it has contributed to the achievement of health sector objectives. If this cannot be shown, at least in the long term, then it is doubtful that the considerable effort and sacrifice required by reform programmes can be justified.

A number of conclusions can be summarised from our discussion of the methods available.

There are three broad approaches to assessing the degree of association between indicators of policy objectives and policy change. The first uses an analytic description of the reform process, often incorporating descriptive indicators which are inherently associated with the policy change. Such a descriptive analysis is essential for any reasonable attribution of causality to be made. The second method adds a longitudinal approach to the descriptive analysis, in which the timing of the policy change is compared with the trend in the indicator concerned. The third uses a cross-sectional approach, comparing observations of the same indicators in settings (e.g.

facilities, districts, etc.) in which the extent of policy implementation varies. Where feasible, it is desirable to combine cross-sectional with longitudinal approaches (in addition to the descriptive analysis) to strengthen confidence in one's conclusions about the changes that are due to the reform.

Second, both longitudinal and cross-sectional approaches encounter the problem of controlling for the influence of external factors. Simple methods can be used to identify the extent to which this problem is likely to exist in any individual situation. These include, for longitudinal approaches, looking for discontinuities in trends where a policy change is quite discrete; looking for changes which are sufficiently large to exclude the likelihood of a long term trend; and generating hypotheses about other explanations of a trend which it is then attempted to reject. For cross-sectional approaches, experimental studies should minimize the expected influence of confounding variables in the design stage. Uncontrolled studies may still have some scope to do this if there is a wide choice of observations to select from, at least on one of the 'intervention' and 'control' sides of the comparison. Failing this, hypotheses regarding alternative explanations of differences between groups can be generated and their viability tested, as with longitudinal approaches. Again, if the nature of the reform (in particular, if it is not national in scope) and the availability of data allow, a time series of cross-sectional data covering periods before and after reform implementation can be very effective in helping the evaluator to separate out the effects of long term and seasonal trends from other possible causes (including the specific reform being evaluated) of observed changes in indicators.

Third, reaching a conclusion as to the effects of a reform in one particular circumstance does not lead automatically to a policy recommendation applicable in other settings or to 'roll out' from local to national implementation.

Finally, the number of potential indicators to measure achievements is great, but there are few indicators which are not liable to misinterpretation or do not require careful and considered use.

If the result of using these methods is that confounding factors cannot be ruled out as the explanation of a change in the indicator, only more sophisticated multi-variate techniques can result in greater certainty about the impact of the policy change. Nevertheless, important and substantial policy impacts are likely to be detectable without this degree of sophistication. Simple methods carefully used, combined with focused descriptive analysis, should be capable of detecting major achievements and drawbacks associated with different types of reform in different countries. Using such accessible and low cost methods, individual countries should be able to tailor their reform programmes by developing those policies which have proved themselves effective and rejecting those which have not done so.

References

Appleby, J., P. Smith, W. Ranade, V. Little, and R. Robinson (1994). "Monitoring managed competition." In Robinson, R. and J. Le Grand (eds.), *Evaluating the NHS reforms*. King's Fund Institute, Policy Journals. New Jersey: Berkshire and Transaction Books.

Baker, J.L. and J. van der Gaag (1993). "Equity in health care and health care financing: evidence from five developing countries." In Van Doorslaer, E., A. Wagstaff, and F. Rutten (eds.), *Equity in the finance and delivery of health care: an international perspective*. Oxford: Oxford University Press.

Barnum, H. and J. Kutzin (1993). *Public hospitals in developing countries: resource use, cost, financing*. Baltimore and London: The Johns Hopkins University Press.

Barros, F.C., J.P. Vaughan, and C.G. Victora (1986). "Why so many Caesarean sections: The need for a further policy change in Brazil." *Health Policy and Planning* 1 (1):19-29.

Bartlett, W. and J. Le Grand (1994). "The performance of trusts." In Robinson, R. and J. Le Grand (eds.), *Evaluating the NHS reforms*. King's Fund Institute, Policy Journals. New Jersey: Berkshire and Transaction Books.

Brudon-Jakobowicz, P., J.-D. Rainhorn, and M.R. Reich (1994). *Indicators for monitoring national drug policies: a practical manual*. WHO/DAP/94.12. Geneva: World Health Organization, Action Programme on Essential Drugs.

Carrin, G., J. Perrot and M. Abrial (1995). "Financial needs and government budget options for the functioning of the health system--a simulation model." Geneva: World Health Organization.

Carter, G.M. and P.G. Ginsburg (1985). *The Medicare case mix index increase. Medical practice changes, aging, and DRG creep*. Report to the Health Care Financing Administration, Rand/UCLA Center for Health Care Financing Policy Research.

Cassels, A. (1995). *Health sector reform: key issues in less developed countries*. Forum on Health Sector Reform, Discussion Paper No. 1, WHO/SHS/NHP/95.4. Geneva: World Health Organization.

Collins, C., A. Green, and D. Hunter (1994). "International transfers of National Health Service reforms: problems and issues." *The Lancet* 344:244-50.

Collins, D., G. Njeru and J. Meme (1996). "Hospital autonomy in Kenya: the experience of Kenyatta National Hospital." Working draft. Boston: Data for Decision Making Project, Harvard School of Public Health.

Cumper, G.C., C. MacCormack and G.J.A. Walker (1985). *Evaluation of health care in Jamaica*. London: London School of Hygiene and Tropical Medicine, Evaluation and Planning Centre for Health Care.

Diop, F., A. Yazbeck and R. Bitran (1995). "The impact of alternative cost recovery schemes on access and equity in Niger." *Health Policy and Planning* 10(3):223-240.

Gertler, P. and J. van der Gaag (1990). *The Willingness to Pay for Medical Care: Evidence from two developing countries*. Baltimore and London: Johns Hopkins University Press.

Gilson, L., P. Kilima and M. Tanner (1994). "Local government decentralization and the health sector in Tanzania." *Public Administration and Development* 14:451-77.

Hanson, K. and L. Gilson (1996) *Cost, Resource Use and Financing Methodology for Basic Health Services: A practical manual (second edition)*, Bamako Initiative Technical Report Series Number 34, New York: UNICEF.

Jamieson, D.T., W.H. Mosley, A.R. Measham, and J.-L. Bobadilla (1993). *Disease control priorities in developing countries*. Oxford Medical Publications (published for the World Bank). New York: Oxford University Press.

Janovsky, K. (1995). "Decentralization and health systems change: a framework for analysis." Revised working document. WHO/SHS/NHP/95.2. Geneva: World Health Organization, Health Systems Development Programme.

Janovsky, K. and A. Cassels (1996). "Health policy and systems research: issues, methods, priorities." In Janovsky, K. (ed.), *Health policy and systems development: an agenda for research*. WHO/SHS/NHP/96.1. Geneva: World Health Organization.

Jolly, R. and M. King (1966). "The organization of health services." In King, M. (ed.), *Medical Care in Developing Countries*. Nairobi: Oxford University Press.

Jolly, R., F. Kamunvi, M. King and P. Sebuliba (1966). "The economy of a district hospital." In King, M. (ed.), *Medical Care in Developing Countries*. Nairobi: Oxford University Press.

Kupor, S.A., Y. Liu, J. Lee and A. Yoshikawa (1995). "The effect of copayments and income on the utilization of medical care by subscribers to Japan's national health insurance system." *International Journal of Health Services* 25(2):295-312.

Kutzin, J. (1995). "Health financing reform: a framework for evaluation." Revised working document. WHO/SHS/NHP/96.2. Geneva: World Health Organization, Health Systems Development Programme.

LaFond, A. (1995). *Sustaining primary health care*. The Save the Children Fund. London: Earthscan Publications Ltd.

Litvack, J. and C. Bodart (1993). "User fees plus quality equals improved access to health care: results of a field experiment in Cameroon." *Social Science and Medicine* 36(11):1397-1405.

Mahapatra, P. and P. Berman (1994). "Using hospital activity indicators to evaluate performance in Andra Pradesh, India." *International Journal of Health Planning and Management* 9(2):199-211.

Maxwell, R.J. (1996). "Hospital autonomy." Geneva: World Health Organization, Health Systems Development Programme.

McPake, B. (1996). "Financing." In Janovsky, K. (ed.), *Health policy and systems development: an agenda for research.* WHO/SHS/NHP/96.1. Geneva: World Health Organization.

Moens, F. (1990). "Design, implementation and evaluation of a community financing scheme for hospital care in developing countries: a pre-paid health plan in the Bwamanda health zone, Zaire." *Social Science and Medicine* 30(12):1319-27.

Newbrander, W., with C. Auma and M. Njau (1995). "Equity and coverage of health care provision in Kenya." Report produced by the BASICS Project for the Health and Human Resources Analysis in Africa (HHRAA) Project, USAID. Washington, DC.

Nittayaramphong, S. and Tangcharoensathien, V. (1994). "Thailand: private health care out of control?" *Health Policy and Planning* 9(1):31-40.

Pabón Lasso, H. (1986). "Evaluating hospital performance through simultaneous application of several indicators." *Bulletin of the Pan American Health Organization* 20(4):341-57.

Prescott, N. (1997). "Setting priorities for government involvement with antiretrovirals." In van Praag, E., S. Fernyak, and A.M. Katz (eds.), *The implications of antiretroviral treatments.* Informal consultation, WHO/ASD/97.2. Geneva: World Health Organization in collaboration with UNAIDS.

Quick, J.D. and S.N. Musau (1994). *Impact of cost sharing in Kenya: 1989-1993.* Boston: Management Sciences for Health.

Rannan-Eliya, R.P. and P. Berman (1993). "National health accounts in developing countries: improving the foundation." Data for Decision Making Publication Number 2. Boston: Harvard School of Public Health.

Ruwe, M., M. Macwan'gi, and S. Atkinson (1996). "Reforming health care provision in Lusaka, Zambia." In Atkinson, S., J. Songsore, and E. Werna (eds.), *Urban health research in developing countries: implications for policy.* Oxford: CAB International.

Simborg, D.W. (1981). "DRG creep: a new hospital-acquired disease." *New England Journal of Medicine* 304(26):1602-1604.

Van Lerberghe, W. and Y. Lafort (1990). "The role of the hospital in the district: delivering or supporting primary health care?" *Current Concerns*, SHS Paper number 2. WHO/SHS/CC/90.2. Geneva: World Health Organization.

Van Doorslaer, E. and A. Wagstaff (1993). "Equity in the finance of health care: methods and findings." In Van Doorslaer, E., A. Wagstaff, and F. Rutten (eds.), *Equity in the finance and delivery of health care: an international perspective*. Oxford: Oxford University Press.

Waddington, C. and K.A. Enyimayew (1990). "A price to pay, part 2: the impact of user charges in the Volta Region of Ghana." *International Journal of Health Planning and Management* 5(4):287-312.

Whitehead, M. (1994). "Is it fair?: Evaluating the equity implications of the NHS reforms." In Robinson, R. and J. Le Grand (eds.), *Evaluating the NHS reforms*. King's Fund Institute, Policy Journals. New Jersey: Berkshire and Transaction Books.

WHO (World Health Organization) (1996). Management indicators for health human resources: developing a menu of indicators. Second stage report (unpublished) prepared by the Centre for Health Planning and Management, Keele University. Division of Organisation and Management of Health Systems. Geneva.

WHO (World Health Organization) (1992). *The hospital in rural and urban districts.* Report of a WHO study group on the functions of hospitals at the first referral level. WHO Technical Report Series 819. Geneva.

World Bank (1994). *Better Health in Africa.* Washington, D.C.

Yang, B.M. (1991). "Health Insurance in Korea: opportunities and challenges." *Health Policy and Planning* 6(2):119-29.

Yang, P., V. Lin, and J. Lawson (1991). "Health policy reform in the People's Republic of China." *International Journal of Health Services* 21(3):481-491.

Yoder, R.A. (1989). "Are people willing and able to pay for health services?" *Social Science and Medicine* 29(1):35-42.

Annex: Indicators of Health Policy Objectives, Ideas on their Interpretation and Possible Data Sources

In this annex, we discuss indicators (and indicator types) which might be used to assess health system performance and measure the effects of reforms on the following objectives:

- allocative efficiency;
- technical efficiency;
- equity in access to care;
- equity in finance; and
- financial sustainability.

After consultation of a selection of literature which evaluates health reform, we produced lists of indicators (and types of indicators) of the effects of changes with respect to each of these objectives. *We are not suggesting that each of these indicators be tracked.* Those who are responsible for monitoring health systems and evaluating reforms must choose the indicators that they deem most appropriate for their context and purposes. Although a large number of indicators are described, they are intended to be illustrative rather than exhaustive. Clearly, not all these indicators relate to all levels and types of change. Although listed by objective at the beginning of each subsection of this annex, the indicators are not all exclusively indicative of a single objective. In particular, there are major overlaps between indicators of allocative and technical efficiency, and of equity of access. Such points of overlap are noted as they arise.

Some of the indicators listed are very specific (e.g. hospital admissions per 100 health centre contacts), whereas others are more general (e.g. distribution of resources among referral levels) and only give an idea of the type of indicator that might be appropriate for measuring progress towards a sectoral objective. In actual practice, the analyst should define indicators very specifically.

Alternative sources of information and data collection methods which might be used to measure these indicators are considered, and the accuracy with which each indicator is likely to reflect the achievement of each objective is discussed. This discussion of the 'accuracy' of indicators is essential because many indicators are ambiguous and require interpretation. In other words, just because an indicator (e.g. unit cost) has gone up or down does not immediately tell us if something positive or negative has happened. Understanding of the specific context in which an indicator has changed is necessary for making an informed interpretation of this change.

A.1 Allocative efficiency

Definition. Allocative efficiency requires that resources are allocated to the activities in which they have the highest value. Using resources for low priority activities which have a small impact on health (such as high technology interventions of dubious efficacy) while higher priority activities which have a major impact on health (such as immunisations) go unfunded is an example of allocative inefficiency.

INDICATORS

Shares of expenditure and resource use: national level
- [a] % expenditure on all hospitals
- [b] % expenditure on tertiary and quaternary hospitals
- [c] distribution of resources among referral levels
- [d] % expenditure on 'public health'

Shares of expenditure and resource use: local level
- [e] shares of expenditure on identified high and low priority activities
- [f] shares of staff time on identified high and low priority activities

Utilization patterns
- [g] level of first contacts and referrals with providers of different levels of care
- [h] hospital admissions per 100 health centre contacts
- [i] relative levels of hospital average length of stay and bed turnover rate
- [j] level of provision of diagnostic equipment

Indicators for health systems with a substantial insurance sector
- [k] utilisation rate of insured relative to uninsured populations
- [l] unit expenditures in insurance facilities relative to MOH facilities
- [m] resource availability in insurance facilities relative to MOH facilities

Assessment of Indicators and Information Sources.

Shares of expenditure and resource use: national level

Some indicators ([a] % expenditure on all hospitals; [b] % expenditure on tertiary and quaternary hospitals; and [c] distribution of resources among referral levels) aim to measure the extent of allocation of expenditure by level of care. The base (source of both numerator and denominator) of these indicators could be total national health expenditure, expenditure from the Ministry of Health budget, expenditure by insurance institutions, or any other source which the health reform is expected to affect. In most countries, allocative efficiency would be improved by the reallocation of total expenditures to lower levels, and this would guide the interpretation of these indicators. These indicators could equally be included under the objective 'equity of access' under the assumption that this too is improved by the reallocation of resources to lower levels.

There are few countries in which an analysis of distribution of total expenditures by level does not indicate substantial allocative inefficiency and inequity of access. However, it does not necessarily follow that reallocation towards lower levels represents improvement. Using the share of expenditures allocated to all hospitals [a] as a measure of allocative efficiency or equity of access implies that all hospital expenditure is low priority. In practice, hospital activity includes supervision of primary health care and sometimes covers outreach clinics which offer high priority activities. The role of the district hospital in quality assurance of activities throughout the district is now well recognised (for example, World Bank 1994; WHO 1992; Van

Lerberghe and Lafort 1990). Reduced expenditure on district hospitals could affect high priority services, and if expenditure reallocated elsewhere is affected (e.g. because the supervision system breaks down), allocative efficiency could worsen. The meaning of this indicator can be quite ambiguous, and, therefore, it should be used with caution.

It may be safer (i.e. less ambiguous) to focus on the share of expenditure allocated to tertiary and quaternary hospitals [b]. In these, high priority activities are certainly fewer. If a poor country seeks to focus public resources on highly cost-effective interventions, the estimates of Jamison *et al.* (1993) suggest that very few interventions offered only by higher level referral hospitals would be included. While some spending on these facilities is necessary, it may be safe to assume that any reallocation of resources from this to other levels of care would reflect an improvement in allocative efficiency. The reason is that, in most countries, public spending on these facilities already absorbs a significant share of total government health expenditures, and most of the services that they provide are, in fact, primary and first referral services that could be more efficiently and appropriately provided in lower cost settings. Nevertheless, choice of this indicator ignores improvements arising from changes in expenditure allocations between other levels of care and interactions between public and private health expenditures. An overall assessment of the distribution of resources among referral levels [c] in conjunction with information about the uses of expenditures will often be the most appropriate course. A related indicator is suggested by Brudon-Jakobowicz, Rainhorn and Reich (1994) in their manual on drug policy indicators: the share of the public drug budget absorbed by 'major' hospitals, which they define as "national and/or regional hospitals" (p.124), but which one could also choose to define according to the national definition of high level hospitals (e.g. tertiary, quartenary, central, teaching, etc.). This indicator measures the distribution of a key health system input, pharmaceuticals, across levels of care, and as such reflects government's resource allocation policies.

An alternative to the share of hospitals in expenditure is a direct estimate of the share of public health in expenditure (indicator [d]). To use this indicator, 'public health' must be defined in such a way that existing budget categories can be used to gain an indication of its share in the total. Sometimes this implies a broad definition. For example, a given level of care alongside any public health related programmed activity financed by the institution in question, such as occupational health or environmental sanitation, might have to be equated to 'public health'. Other budgeting systems may allow a more detailed breakdown of expenditures within levels of care. Clearly, the more specific the breakdown possible, the more meaningful the indicator.

Data on total national health expenditure are not usually readily available, unless a specific study has been done, since it requires some estimate of the size and distribution of private sector expenditures not only in large private and mission hospitals but in small private clinics and the traditional and informal sectors. Estimation of this is unlikely to be within the scope of small scale studies, since household surveys are almost the only way to get comprehensive information about such a wide range of sources of care. An existing study of the extent and nature of private expenditure, even if only carried out once, could indicate the likely trend in distribution of total expenditure if it is thought that a reform has changed the balance between expenditures in the public and private sectors, or between insurance-covered service provision and unorganised individual use of health services. Often, use can be made of general household budget or

expenditure surveys (i.e. information that was not specifically intended to examine the health sector) implemented by national statistical bureaus. Such surveys exist in many countries and do contain useful, though limited, information that can be used to generate rough estimates of private health care spending.

Shares of expenditure and resource use: local level

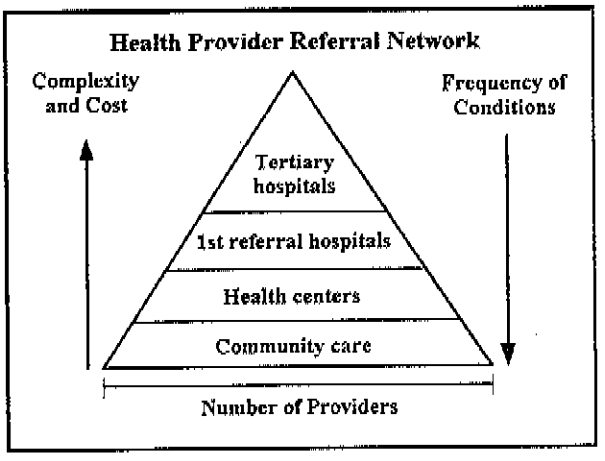
At local level, for example within a specific facility or district, indicators of allocative efficiency can be more specific ([e] shares of expenditure on identified high and low priority activities; and [f] shares of staff time on identified high and low priority activities). These are unlikely to mislead but are not sensitive to changes in utilisation patterns between facilities, levels of care or sectors. Indicators [e] and [f] may also only imperfectly reflect changes in utilisation patterns of the facility itself--for example from more to less serious cases, or from curative to preventive service use.

Data to measure [e] and [f] could be derived from a sample of facilities and the use of the "Cost, Resource Use and Financing Methodology" (Hanson and Gilson 1996) or similar.

Utilisation patterns

In a hierarchical referral system, there are defined providers of first contact (health centres or posts), first referral (e.g. district hospitals), and tertiary (e.g. central hospitals) services. Reforms intended to improve the performance of the referral system are often oriented towards increasing the use of health centres for primary care and hospitals for referral care. Comparison of the levels of first contacts with different types of providers [g] gives a sense whether the pattern of service provision and use is efficient. The extent to which first referral and tertiary hospitals are used to provide first contact care is an indicator of allocative inefficiency because (1) it is likely that such care could be provided less expensively in a lower cost setting, and (2) the use of staff trained to undertake complex diagnostic and therapeutic procedures to provide basic care reflects an inefficient use of the health system's human resources. Similarly, if data are available on referral rates, comparisons across hospitals or of the same hospital over time could indicate differences in performance. If no data are available on actual numbers of cases referred to hospitals, it may be possible to construct a proxy, such as hospital admissions per 100 health centre contacts [h].

Figure A-1. A model of the economic rationale for a hierarchical referral system



The starting point for analyzing the allocative efficiency of a regional or national system of hospitals is a conceptual understanding of the role that hospitals of different levels of complexity are supposed to play in the health system. As suggested by Figure A-1, a referral system is conceptualized as providing the most cost-effective distribution of services based on the costs of care and population epidemiology and distribution. Although there are some exceptions, truly severe conditions requiring sophisticated tertiary care occur relatively infrequently, whereas more routine conditions

occur more frequently. National tertiary hospitals are designed to diagnose and treat highly complex cases and thus have an intensive concentration of skilled staff and medical equipment. Most countries have a type of hospital that is of intermediate complexity between national tertiary hospitals and district hospitals. The number and location of all hospitals of varying levels of complexity should be determined by conditions of cost-effectiveness related to resource constraints, epidemiology and population distribution.

At the hospital levels of the health system, some inpatient service statistics ([i] average length of stay (ALOS), and bed turnover rate; see Box A-1) can be used as indicators of whether the pattern of patient management across different types of hospitals is appropriate. If the system is working as intended and case mix is appropriate to the type of hospital, district hospitals should have a higher turnover and shorter ALOS than tertiary hospitals, with the values of these indicators for intermediate ("2nd referral") hospitals falling in between. By comparing the actual pattern of inpatient service statistics to this pattern (shown in Table 1), it may be possible to identify some problems related to the allocative efficiency of the health system. For example, if the central tertiary hospitals have the highest turnover rates, this may mean that they are seeing many patients who could be treated in a less complex, lower cost facility.

Box A-1. Inpatient Service Statistics

- average length of stay (ALOS): average number of days per inpatient stay per patient
- bed turnover rate: the annual average number of patients per hospital bed, a measure of productivity

$$\text{ALOS} = \frac{\text{inpatient-days}}{\text{admissions (or discharges + deaths)}}$$

$$\text{turnover rate} = \frac{\text{admissions (or discharges+deaths)}}{\text{beds}}$$

Table A-1. Desired Relative Pattern of Inpatient Service Statistics

Hospital type	Average length of stay	Turnover rate
1st referral	Lower	Higher
2nd referral	Middle	Middle
Tertiary	Higher	Lower

Indicators [g] to [i] may be produced from routinely collected facility data, which may be available (depending on the specifics of each country's information management system) at national, regional, local, or facility level. Facility-based and household surveys can also be used to generate information on referral patterns.

In some (often middle income) countries, it may be useful to assess trends or make comparisons of levels of provision and utilisation of diagnostic equipment (j). International comparisons of the number of Computerized Tomographic (CT) scanners per million population were used in a Thai study to show how high-cost medical technologies have proliferated in Bangkok, for example (Nittayaramphong and Tangcharoensathien 1994). Such comparisons can facilitate policy-relevant conclusions even though it is not easy to identify the appropriate level of equipment provision in absolute terms. If rapid growth is noted following a reform or a major

difference is revealed between insured and uninsured patients or facilities, for example, there is likely to be cause for concern. Even if appropriate levels of provision have not yet been surpassed, the experience of many countries suggests that such cost elements have a tendency to evade control, eventually representing serious divergence from allocative efficiency. Some countries have data at national level regarding provision of diagnostic equipment, and identification of numbers of specific items available at lower level such as facility or district is a simple matter.

Indicators for health systems with a substantial insurance sector

Most of these indicators ([k] utilisation rate of insured relative to uninsured populations; [l] unit expenditures in insurance facilities relative to MOH facilities; and [m] resource availability in insurance facilities relative to MOH facilities)[2] measure relative expenditure and resource concentrations in insured and uninsured populations and facilities. These are likely to reflect both allocative efficiency and equity of access (since in most countries insured populations are both richer and in better health).

However, these indicators are quite difficult to interpret. The introduction, expansion or reform of insurance may add to the total revenues available to the health sector, and an increasing proportion of resources allocated to the insured does not preclude an increasing total resource allocation to the uninsured. It is often argued (but has rarely been demonstrated) that provision of separate facilities for the insured or expanding the number of formal sector workers covered by health insurance leads to better targeting of public health expenditures.

Data relating to the insured population, expenditures of insurance funds and resource availability in insurance facilities can often be obtained from the routinely collected data of the insurance funds themselves. With data on public sector expenditure and facility utilisation, some comparisons between the insured and the uninsured can be made quite easily (e.g. per capita expenditure on behalf if insured and uninsured populations, per capita utilisation, etc.). Utilisation and facility-based data may require specific studies in a sample of public facilities. Care must always be taken to establish the extent of comparability of data taken from different sources by understanding and documenting the methods of compilation and collection of data.

A.2 Technical efficiency

Definition. Technical efficiency requires that once the activity or output mix has been determined (for example the mix of curative and preventive services to be offered), the activities are carried out without wastage of inputs and at the minimum possible cost.

---

[2] Of course, indicators [l] and [m] are only relevant in countries that have separate provider facilities for insured patients.

Assessment of Indicators and Information Sources. Usually, the most important inputs for health service delivery are salaries and drugs. Therefore indicators of appropriate drug use and staffing ([a] to [g]) are important indicators of technical efficiency.

INDICATORS

Indicators of efficient drug use and management

[a] number of drugs per prescription
[b] % of prescriptions that include injections
[c] % of under-fives with diarrhoea receiving ORS relative to anti-diarrhoeals
[d] average stockout duration at various storage levels of the health system

Indicator of efficient staff use

[e] % staff in total recurrent expenditure
[f] utilisation per provider
[g] absenteeism rate

Facility based indicators

[h] average cost of any unit of activity (e.g., curative visits to health centres, inpatient days, prescriptions, lab tests)
[i] admission or discharge rates
[j] inpatient service statistics (bed occupancy rates, bed turnover rates, and average length of stay)

Indicators of efficient drug use and management

Reforms to systems and practices for procuring, financing, managing, and prescribing drugs are a necessary part of broader reforms to the overall health system in most countries. The WHO manual on *Indicators for Monitoring National Drug Policies* (Brudon-Jakobowicz, Rainhorn and Reich 1994) provides a wide range of indicators to assess the structure, process, and outcomes of national drug policies and practices, and is an essential reference for those who are responsible for analyzing reforms to drug systems. We draw on a small number of these indicators to illustrate how they might reflect changes in the technical efficiency (and quality) of the entire health system in response to a policy change (e.g. implementation of drug user fees combined with prescribing protocols).

The first three of these indicators ([a] number of drugs per prescription, [b] % of prescriptions that include injections,[3] and [c] % of under-fives with diarrhoea receiving oral rehydration salts (ORS) relative to anti-diarrhoeals) are meant to reflect the extent to which prescribing and dispensing practices are *rational* (i.e. appropriately used, from a qualitative perspective). In general, improvements in quality through more rational prescribing practices will be synonymous with greater technical efficiency and lower cost per patient. While there is no global standard for the 'correct' average number of drugs per prescription[4] or percent which should include injections, most countries suffer from excessive use. Moreover, indicator [c] can be compared to a standard, which is that ORS should always be used and anti-diarrhoeals never used. To assess the effect of reforms, these indicators can be tracked for the same groups of providers over

[3] For the purpose of calculating this indicator, immunizations should not be counted as injections.

[4] Brudon-Jacobowicz, Rainhorn and Reich (1994) suggest, however, that there is a consensus "that an average of more than two drugs per prescription will probably reflect a problem in prescribing practices" (p.155).

time (longitudinal analysis) or for different groups of providers (e.g. public vs. private, insured vs. uninsured, urban vs. rural, etc.) at the same time (cross-sectional analysis).

Indicators related to stockouts ([d] average stockout duration at various storage levels of the health system) may suggest a number of factors that relate to the efficiency of drug management, since several factors could cause drugs to be out of stock. These might include any (or some combination) of the following: poor stock management, procurement problems, delays in distribution, inadequate funding, or unanticipated high demand. Thus, if the extent of stockouts is observed to change over time or vary across locations or sectors, a more detailed analysis of the exact causes of change will be needed.

The prescribing indicators ([a], [b], and [c]) indicators apply to outpatients only, and the data needed to calculate them can be derived from low cost surveys of records or observations of practices at private and public drug outlets. The WHO manual provides detailed guidance for sampling public and private drug outlets and examples of data collection forms.

Stockouts [d] can be measured at intermediate levels of the system, such as central or regional medical stores, or at final distribution/sales points, such as public or private drug outlets. The WHO manual suggests calculating stockouts as the percentage of time that individual items from a defined 'basket' of drugs are out of stock, relative to the average stockout duration for the same basket during the previous three years. However, the calculation of a stockout indicator may need to be modified according to the data available in a particular setting or at a particular point in time. As with the drug use indicators, the manual provides detailed guidance on how to gather and organize the data needed to calculate this indicator.

Indicators of efficient staff use

The performance of health care workers is directly and indirectly affected by health system reform processes. As with the others described in this document, any single indicator of the staff efficiency can be more accurately interpreted in the context of other indicators and information on working conditions and the appropriateness of staff mix at various levels of the health system. It is therefore best to use a set of indicators rather than any single measure to assess the extent to which the health workforce is 'well-managed' (WHO 1996).

Reform often affects the total health care budget devoted to staff. A reduction in indicator [e] (the percentage of staff costs in the total health system or health facility budget) could be regarded as a technical efficiency gain if quality is maintained at a corresponding acceptable level. Indicators that link the work of health care providers to the delivery of services [f] can also provide useful information on the impact of reforms. This indicator might be calculated as the number of patient contacts (inpatient plus outpatient) per total number of skilled staff persons, per doctor, or per nursing person.

The number of days absent as a proportion of total working days by staff category [g] also provides an indication of the effect of a reform on health worker and system performance. If it is assumed that an efficient and productive system will have a lower rate of absenteeism, then this indicator would reflect the efficiency of management structures within the system, as well as the motivational levels of health workers.

Data for these indicators should be readily available from health system and facility records, at least for government health facilities and services. For [e], it is better to use data on actual expenditures rather than amounts budgeted at the beginning of the fiscal year. It is likely that [f] would be calculated at a number of facilities rather than for the system as a whole, enabling comparisons of performance across regions or sectors (or of the same facility or group of facilities over time). One must be careful to confirm staffing level data with periodic facility checks, however, because the reported figures may not always match the numbers and categories of staff present and active, especially if there is poor documentation of unfilled vacancies, of locally employed supplementary staff, or of chronic absenteeism. Data on absenteeism [g] would most probably require facility visits, a detailed review of their records, and direct observation to compare actual with reported staffing levels.

Facility-based indicators

Average (or unit) cost comparisons ([h]) might be used to assess technical efficiency but there are important limitations which indicate the need for caution. For a given mix and quality of services, a lower unit cost implies better technical efficiency. However, in those countries in which health services are believed to be under funded, increases in unit cost (which might arise because drugs are now available when not before, for example) may indicate technical efficiency improvement (as a result of quality improvement). In other countries in which cost containment rather than under funding is the major concern, such indicators may be more straightforward. In both cases, ensuring that 'like with like' is being compared should ensure that a comparison of unit cost is more meaningful. In these circumstances, higher unit costs may be explained by under utilised staff, equipment or facilities which might not be noticed by the other indicators discussed. In circumstances where unit cost differences are explained by drug use or diagnostic equipment use differences, it is better to measure this directly, giving more opportunity to consider what the appropriate levels might be (see Barnum and Kutzin 1993, pp.85-88, for a discussion of difficulties in interpretation of unit cost findings).

Indicators [i] (admission or discharge rates) and [j] (inpatient service statistics: bed occupancy rates, bed turnover rates, and average length of stay) are specifically for use in assessing changes at hospital level. These indicators must be used in conjunction with assessment of quality of care before they can be assumed to reflect efficiency. Where a policy change might have resulted in different admission or referral criteria or a different mix of inpatients (e.g. as a result of increasing use of day surgery), it is important also to ensure that hospital "case mix" (the complexity of cases presenting at a hospital) is monitored sufficiently to pick up any changes in overall severity and complications which might explain changes identified. If it is not deemed feasible to measure case mix quantitatively, a qualitative assessment should be made so that this factor is considered before drawing conclusions from the inpatient indicators.

Three basic inpatient service statistics [j] hospital bed occupancy rates, bed turnover rates, and average length of stay can be used together to give an indication of the relative efficiency of a hospital's inpatient utilisation compared to other, similar types of hospitals, or to itself over time (Pabón Lasso 1986). Recalling the formulae for bed turnover rate and ALOS from Box A-1, and considering also the formula for the bed occupancy rate in Box A-2, it is clear that the only data needed to calculate all of these statistics for a hospital are (1) number of beds, (2) number of admissions or discharges, and (3) number of patient-days. These basic data and the service

statistics are generally available at the hospital level, but in many countries some or all of the service statistics are reported to regional or national levels. Most health systems and studies focus solely on the occupancy rate, but this can be misleading without also having information on the bed turnover rate and average length of stay. For example, a high occupancy rate might be interpreted as a sign of efficiency when, in fact, it might mean that patients are having unnecessarily long lengths of stay. Thus, a more accurate interpretation of these indicators requires that they be used together. As the formulae in the lower part of Box A-2 illustrate, these three indicators are mathematically related, so that knowledge of any two enables the third to be calculated. Using this knowledge and hospital data from Colombia, Pabón Lasso (1986) developed a graphic technique to present the three statistics together and guide their joint interpretation. This has been applied in analyses of hospital performance in Indonesia (Barnum and Kutzin 1993), India (Mahapatra and Berman 1994), and elsewhere.

Box A-2. More on Service Statistics

- % bed occupancy (occupancy rate): the percent of beds filled with patients for a defined period of time (e.g., one year), a measure of capacity utilisation

$$\%\ \text{bed occupancy} = \frac{\text{inpatient-days}}{365 * \text{beds}}$$

$$\text{Occupancy} = \frac{\text{ALOS} * \text{Turnover}}{365}$$

$$\text{Turnover} = \frac{\text{Occupancy} * 365}{\text{ALOS}}$$

$$\text{ALOS} = \frac{\text{Occupancy} * 365}{\text{Turnover}}$$

Care must always be taken in interpreting these indicators, especially when drawing conclusions about efficiency, for the same reasons that caution is warranted for the interpretation of unit cost data. By themselves, these indicators do not incorporate variations in the quality of services being provided, and they also may reflect differences in the mix of patients being seen in particular facilities rather than the efficiency with which a hospital manages its patient load. For comparisons of performance between hospitals, it is important to try and correct for variations in case mix. A reasonable, though imperfect, way of doing this is to only compare hospitals with similar functions in the health system. Thus, for example, one should compare performance measures from one district hospital with that of other district hospitals, not with a mix of hospitals that include central tertiary facilities.

A.3 Equity in access

Definition. The objective of achieving equity in access implies that the *opportunity* to benefit from health care services is distributed according to need rather than according to other factors such as income, insurance status, geographic location, gender, or age. Because this opportunity is difficult to observe, indicators focus on differences between population groups in the receipt/use of health services as a proxy for differences in access.

Assessment of Indicators and Information Sources.

INDICATORS

Utilisation indicators
- [a] utilisation by facility
- [b] utilisation of services by specific population groups
- [c] urban/rural differentials in utilisation
- [d] source of treatment before death by income or demographic groups

Exemption and concession indicators
- [e] % exemptions, price reductions, debts allowed and free insurance entitlements
- [f] correlation between [e] and poverty or income levels by geographical area
- [g] social and economic characteristics of those receiving concessions

Indicators for health systems with an insurance sector
- [h] insurance coverage levels of disadvantaged groups
- [i] economic, social and geographical characteristics of insured and uninsured populations
- [j] total membership of locally based health insurance schemes

Affordability and availability indicators
- [k] price comparisons between public and other sectors
- [l] distribution of providers and facilities

Utilisation indicators

Utilisation indicators ([a] utilisation by facility; [b] utilisation of services by specific population groups; [c] urban/rural differentials in utilisation and [d] source of treatment before death by income or demographic groups) have been used most to assess equity of utilisation and access to care for disadvantaged groups. These indicators might take the form of quantities of services used (e.g. annual or monthly outpatient visits or inpatient admissions) or population-based rates of utilisation (e.g. outpatient visits per capita or inpatient admissions per 100 population). Use of aggregate utilisation data [a] for evaluation of equity assumes that utilisation by the poorest groups is most sensitive to policy change. Household survey evidence from several countries at different levels of economic development on the impact of user fee and quality changes suggests that this common-sense assumption is justified (see, for example, Gertler and van der Gaag 1993; Litvack and Bodart 1993; Kupor *et al.* 1995).

The problem with aggregate utilisation as an indicator is that it may hide offsetting utilisation changes. For example, the introduction of a user charge and simultaneous improvement in quality might increase the utilisation of those who had previously used more expensive providers and still reduce the utilisation of those who had previously been unable to afford those providers and are now unable to afford the user fee. Overall utilisation levels could be maintained, and no problem therefore detected. Alternatively, utilisation levels in one facility may be observed to change, but offsetting utilisation changes in other facilities may not be detected. Disaggregated indicators ([b] to [d]) are therefore preferred where their measurement is feasible.

A second limitation which applies to all utilisation indicators is the possibility that a reform may well be accompanied by other changes, such as changes in accounting and recording procedures and training of staff, which could affect the quality of record keeping at a facility. Some reforms, especially those that relate payment of providers to some form of reported performance, also generate incentives for more complete recording of utilisation. This occurred in the United States

after the social health insurance scheme for the elderly (Medicare) introduced a system of paying hospitals on the basis of reported diagnostic and treatment information (Carter and Ginsburg 1985; Simborg 1981). Thus, the utilisation indicators may reflect completeness of recording rather than actual utilisation levels. It is important to consider and, if necessary, investigate, whether changes observed may be explained in this way.

Data for aggregate utilisation can be obtained from routine data collection systems (at least for government facilities). Disaggregated data [b] presents difficulties according to what type of disaggregation is attempted. There is not usually any explicit information regarding economic status which is routinely collected; therefore, information relating service use to individual or household income levels requires survey data (see Baker and van der Gaag (1993) for analyses of utilisation by income group in Bolivia, Côte d'Ivoire, Ghana, Jamaica, and Peru). Data on utilisation by women and by under-5s and/or under-1s usually are available in regular reporting systems, and may suggest the degree to which overall trends reflect disaggregated trends, as well as the utilisation of these priority groups. If it is possible to identify the location of patients' residence, it may be possible to infer the utilisation of different economic groups from this. However, distance plays an independent role in households' health care seeking behaviour, and this analysis will be more conclusive if areas of differing economic status but similar distance can be compared. Conducting occasional surveys of inpatient and outpatient origin in health facilities can be a low cost way of gathering information on the relation between the geographic location of households and utilisation patterns. This involves asking a sample of patients where they live or how much time it took them to travel to the health facility (see Cumper, MacCormack and Walker (1985) for an example from Jamaica, plus Van Lerberghe and Lafort (1990), Jolly and King (1966) and Jolly *et al.* (1966) for discussions of the uses of this type of analysis).

Comparisons of per capita utilisation trends in rural and urban facilities [c] enable an assessment of whether geographical inequities are widening or narrowing in response to reforms. This indicator does not reflect the extent to which rural populations use urban facilities, however (a factor which could be influenced by a reform which increases adherence to the referral system, for example). Again, a patient origin survey can help with the interpretation of these indicators.

Indicator [d] (source of treatment before death by income or demographic groups) suggests an alternative to facility-based data for the assessment of disaggregated utilisation trends. If deaths are registered adequately, a study can be undertaken based on the population included in the register of deaths and this indicator measured. Such a study would be much smaller than a general household health service utilisation study, could control for severity of condition and enable detailed information on household characteristics to be collected.

Exemption and concession indicators

The indicators [e] (% exemptions, debts allowed and free insurance entitlements) measure the extent to which policies in place to address access to care for low income and other priority groups are being implemented. These are likely to be available from facility or national records. If they are assumed to reflect equity of access, it has also to be assumed that they are targeted appropriately, or at least more often targeted appropriately than not. One rough test of this is implied by indicator [f] (correlation between [e] and poverty or income levels by geographical

area). If targeting is appropriate, the geographical distribution of exemptions, free insurance coverage and other measures would be expected to reflect the distribution of poverty or income. Nevertheless, it is quite possible that geographical targeting is adequate, whereas its application to individuals still works perversely, favouring civil servants and the more politically powerful, for example. Direct study of the characteristics of those exempted from charges or offered reductions or loans is required before indicator [g] (social and economic characteristics of those receiving concessions) can be measured. Examples of possible data sources include exit polls at facilities or follow up of households who are recorded as having received exemption or other concession (for an illustration from Kenya, see Newbrander 1995).

Indicators for health systems with insurance or prepayment schemes

There are some indicators which specifically relate to evaluation of insurance reforms ([h] to [j]). Coverage levels of disadvantaged groups [h] give a partial picture of their access to services, although it is often the case that despite insurance coverage, poor physical access to services results in continued low utilisation or utilisation of inferior quality services. The same limitation applies to indicator [i] (economic, social and geographical characteristics of insured and uninsured populations), since across groups which are disadvantaged for these reasons, there are other barriers to access than financial ones. Still, in most cases, it is reasonable to assume that the insured have greater access than the uninsured.

As stated in section A.1, allocative efficiency indicators [k], [l] and [m] are also relevant indicators for examining equity in utilisation. By comparing utilisation rates between insured and uninsured persons (indicator [k] from A.1), an assessment can be made of the impact of insurance on equity of access to care. The interpretation of this indicator may differ in situations where there are compulsory (social) health insurance systems, as compared to voluntary arrangements. Where insurance coverage is voluntary, there is always the chance that the observed utilisation by insured persons reflects their health status rather than their insurance status. In other words, they may have obtained insurance coverage because they knew that they were likely to use care ('adverse selection'). In this context, this pattern of utilisation is more of an indicator of the financial (un)sustainability of insurance schemes than of the equity consequences of expanded insurance coverage.

Since locally based health insurance schemes are usually intended to expand utilisation of services in a specific locality which has previously been characterised by low utilisation, and thus to improve equity of access in relation to other areas, the total membership of such a scheme [j] is an indicator of its success in this respect.

Other indicators of affordability and availability of health services

Price comparison of public and other sectors [k] enables the concept of 'relative affordability' to be evaluated. Since disadvantaged groups are most likely to seek care where it is cheapest, the relative affordability of the public sector determines their access to that sector. Still, interpretation of the indicator relies on an assessment of the quality of public sector services in relation to competing alternatives. If the public sector aims to provide at least some minimum standard of quality to the poorest groups, its undercutting of sources which are thought not to provide at least this quality, for example unlicensed drug sellers and practitioners, is an indicator of its achievement of this objective. Since the public sector is the sector which the poorest

groups are most likely to use, this indicator is also therefore likely to reflect equity in finance. Although prices of services in the informal sector are not usually routinely collected, they can quite easily be established through interviews with members of the community or through informal approaches to the providers themselves.

The price or cost of specific health services can also be compared with a measure of household expenditures to give a sense of the affordability of services. This can be a powerful measure of equity of access if use can be made of preexisting surveys of household income or expenditure patterns that allow the analysis to be disaggregated into income or expenditure groups. For example, an analysis of the affordability of different options for anti-retroviral therapy for HIV+ persons (Prescott 1997) used data from Thailand to compare the cost of each option with the average nonfood expenditure per capita, disaggregated by decile. Assuming that the average nonfood expenditure represented a realistic budget constraint on individual ability to pay, this comparison showed that a proposed subsidized price of anti-retrovirals would be unaffordable for the poorer 50 percent of the population. A cruder assessment of affordability can be made by comparing the price/cost of a specific treatment to average per capita income if no disaggregated data are available.

The WHO manual on drug policy indicators (Brudon-Jakobowicz, Rainhorn and Reich 1994) identifies some specific price comparisons (with details on data sources and calculation methods) that can be made to assess the affordability of pharmaceuticals. For example, to see if competitive pressures are leading to lower prices in the private sector, perhaps through greater use of generics, the percentage of items in a defined 'basket' of essential drugs that are available at the cheapest price can be determined (p.150). As a way of assessing the affordability of a standardized course of treatment, the manual also suggests measuring the standard retail price of pneumonia treatment and comparing it to the price of a standard daily basket of food (p.151).

Only large scale reforms at national, or at least regional, level are likely to affect the distribution of providers and facilities [l]. For example, this indicator could be used to assess the success of attempts to expand insurance programmes to rural areas in achieving improved physical access to services for rural populations through increasing the demand for services there and attracting new providers. In most countries, registration records will enable the mapping of registered practitioners. Failing this, or if unregistered practitioners are also considered to be important, maps can be constructed for sampled areas through interviews with members of the community and physical search. Brudon-Jakobowicz, Rainhorn and Reich (1994, p.149) suggest measuring the percentage of essential drugs that are (or are not) available on any given day in a selection of health facilities located in remote parts of the country as an indicator of the availability of health services.

A.4 Equity in finance

Definition. Equity in finance implies that contributions for health services are related to ability to pay rather than to health status. This concept can be combined with equity in utilisation to derive a condition for overall (net) equity, which requires that subsidies should flow from the rich to the poor and not from the poor to the rich.

INDICATORS

Breakdown of sources of finance

[a] distribution of taxation burden
[b] breakdown of sources of finance for insurance programmes by government, employer, employee or individual
[c] breakdown of sources of finance for individual facilities by government, insurance, direct payment and other
[d] breakdown of sources of finance for the health sector overall

Indicators of distribution of expenditure

[e] geographical distribution of government health expenditures
[f] sectoral distribution of government health spending (between direct public sector expenditures, and support to insurance and private sectors)
[g] geographical distribution of insurance agencies' expenditure
[h] distribution of utilisation of different sectors by socio-economic group

Indicators of cross-subsidy

[i] extent of cross-subsidy within a facility through stepped charges (e.g.,. private wards)
[j] extent of cross-subsidy between different insurance funds
[k] comparison of contributions, income levels and benefits across different insurance funds

Assessment of Indicators and Information Sources.

Overview

In a way, assessing equity in finance means answering the question, "who pays?", while assessing equity in utilisation answers, "who receives?". Thus, subsidy flow is determined by the distribution of the financial burden imposed by health services, and the distribution of the utilisation of services. Therefore, equity in finance is closely linked to equity in utilisation, and it usually makes sense to consider the two together. Since finances flow through a number of institutions (including the exchequer, insurance agencies and facilities themselves), determining the distribution of the financial burden involves a breakdown of the sources of finance for each of these (indicators [a] distribution of taxation burden; [b] breakdown of sources of finance for insurance funds/schemes/programmes; [c] breakdown of sources of finance for individual facilities and [d] breakdown of sources of finance for the health sector overall). Each of these institutions then finances a number of different sectors and provider institutions, and the distribution of the expenditures of each has to be analysed separately ([e] geographical distribution of government health expenditures; [f] sectoral distribution of government health expenditures and [g] geographical distribution of insurance agencies' expenditures). Utilisation patterns of each facility or sector have to be analysed ([h]), and even within institutions and sectors, cross-subsidies may take place, for example between private and general wards [i] or between social health insurance funds [j] resulting in inequities when funds are compared [k]. These have to be assessed through analysis of the distribution of financial burden and utilisation in facility departments. Therefore each of

the indicators listed above are components of the information needed for a complete analysis of both equity in finance and equity in access, which in turn can lead to a clearer understanding of the overall equity in the health system and how reforms are affecting this total picture.

Breakdown of sources of finance

Central government accounts provide a certain amount of the information needed to assess the distribution of the taxation burden [a]. They should provide information on proportionate shares of each different type of tax and for income tax, the distribution of the burden between income bands. Analysis of the distribution of the burden imposed by sales and import tax requires separate study, however, which may be considered beyond the scope of a health financing study. In some countries an analysis of the distribution of the burden of taxation may already have been undertaken (e.g. by the Ministry of Finance or Central Bank).

Insurance agencies' own records should enable a breakdown of sources of finance (who's covered, who's contributing) for insurance programmes [b]. Nevertheless, there are ambiguities in separating the employer and employee contribution. It is not clear to what extent voluntary and compulsory employer contributions are offset against wages, and without analysis of the labour market, and even the macro-economic effects of contributions, these ambiguities can only be noted. This analysis should also identify and if possible, quantify, the extent to which government directly (through explicit subsidies) or indirectly (through tax deductions or credits) subsidizes participation in insurance schemes.

For individual facilities [c], the contribution of different sources of finance is not usually known and requires separate, facility-based study. Hanson and Gilson (1996) offer a set of simple guidelines which could be followed. For a comprehensive assessment of the sources of finance for the whole health sector [d], a large household survey and wide ranging facility assessments are probably required. More modest studies might simply combine the measurement of indicators [a] and [b] with some sample facility assessments, noting which health sub-sectors are unlikely to have been properly accounted for in this process.

Evaluating the distribution of the burden of financing from each of these sources can be done using a variety of methods. For example, analyses of equity in finance in 10 OECD countries (summarized in Van Doorslaer and Wagstaff 1993) used sophisticated calculations of indices of the progressivity of each source of funds. Less complex approaches that combine informed judgments with good descriptive information on sources of funds can also be used (Baker and van der Gaag 1993). National Health Accounts can be a useful input into this analysis (Rannan-Eliya and Berman 1993).

Indicators of distribution of expenditure

The geographical distribution of both government [e] and insurance agencies' [g] expenditures relates more to equity in the receipt of subsidies or services than to equity in finance per se. Analysis of this distribution may be straightforward if a decentralised budgetary system is used by which funds are first allocated to regions or districts. In more centralised systems, allocations to individual facilities will have to be collated by geographical areas, and budgetary headings such as personnel or capital equipment will have to be apportioned according to information about the distribution of these items in as much detail as practicable. How easy it is to

disentangle government expenditures on different sectors (direct expenditure on the public sector and support to insurance and private sectors) [f] depends on record keeping systems. Imputing the value of tax concessions will in most cases be feasible, at least approximately, using tax records. The work involved in reaching greater degrees of precision than, for example, assuming all those whose insurance contributions are tax exempt have the average level of income and average insurance package for that group, may or may not be deemed worthwhile, depending on the record keeping system. Final receipt of subsidy and therefore distribution of expenditure between different socioeconomic groups depends ultimately on utilisation [h], measurement of which has been discussed in section A.3 above.

Indicators of cross-subsidy

The extent of cross-subsidy within facilities [i] requires facility level studies in which the costs of services provided within departments and the direct payments received for the services offered by those departments are compared. Such an analysis would be needed, for example, to determine if the prices charged to patients using private beds in public hospitals are fully covering the costs of these beds, or if government is, in fact, subsidizing these 'private' services. Where only a proportion of costs are covered through direct payment in the facility as a whole, this indicator will instead permit a sub-facility analysis of the channelling of outside subsidies.

Cross-subsidy takes place between insurance funds [j] when they purchase services from the same facilities and either negotiate different prices, or predominantly utilise different services between which there is some cross-subsidy. Both facilities' and insurance agencies' accounts should identify prices paid. Facility level analysis of costs and income by department (as for [i]) would enable analysis of the second type of cross-subsidy.

Indicator [k] will normally be capable of assessment through a comparison of the contribution and benefit packages offered by alternative funds, perhaps combined with quality and cost assessment of the benefits offered, and further investigation of the income levels of the population covered as discussed above. Comparisons could also be made with the uninsured and the implicit benefit package available in the public sector. Where contributions and benefits are not proportionate in comparisons of insurance funds, previous indicators may have identified a disproportionate subsidy as the cause. Other causes, such as differences in efficiency between funds or excessive profit in some funds, may reflect issues other than equity of finance. Where there is an association with income level of those covered, some combination of relevance to equity of finance and equity of access is implied. This might be an issue where a separate insurance fund is set up to cover poorer or rural groups and cannot take advantage of the economies of scale available to larger urban-based funds. Identification of consequent inequity of finance or utilisation might suggest a rationale for targeting of public subsidies towards such funds.

A.5 Financial sustainability

INDICATORS

[a] relative growth rates of health sector expenditure, health sector price index, GNP and specific incomes of main paying groups (government, insurance agencies and direct payers)

[b] trends in the percent of funding coming from donors, government, individuals and other sources

[c] cost recovery ratios

Definition. LaFond (1995, p.17) defines sustainability "as the capacity of the health system to function effectively over time with minimum external input." Relatedly, financial sustainability can be narrowly defined as the extent to which national or local health expenditures are funded from national resources, or more flexibly, as the medium to long term stability of a mix of funding sources. An important feature of this definition is that it applies to the health system rather than individual facilities or programmes. While it may sometimes be appropriate to examine sustainability from the perspective of the latter, it is important to remember that, in most cases, financial resources are fungible, especially over a period of time. Thus, an individual health programme or facility may appear to be improving its sustainability when it may, in fact, be drawing resources from the rest of the health system.

Assessment of Indicators and Information Sources. The longer term capacity of nationally based resources to maintain their role in financing the health sector might be assessed using the types of indicators suggested by [a] alongside more qualitative approaches. If information is available on sources and uses of funds in the health sector, the analysis can be facilitated by disaggregating health expenditure into several categories, such as government health expenditures on government health services, donor expenditures on government health services, household expenditures for government health services, total expenditure on government health services, government expenditures on privately provided health services, private expenditures on private health services, etc. Ratios can then be established between some categories of these expenditures and the incomes of these paying groups. Assessments can then be made of the likely capacities of the different groups (e.g. government, private households) to increase their contributions. Disproportionate growth in health sector expenditure relative to GNP and/or the incomes of paying groups would suggest an impending problem of cost containment, which in turn would pose a severe threat to financial sustainability. To assist decision-makers to think through these issues, a simulation model (Carrin, Perrot and Abrial 1995) has been developed as a tool for assessing the feasibility and financial implications of alternatives. If no health sector price index has been calculated, a consideration of the growth rate of the salary bill and the pharmaceutical import bill (unless the country concerned is a major manufacturer of pharmaceuticals) should give an indication of the relationship between health sector and general inflation.

Documenting the patterns and trends of donor, government, and private finance [b] may suggest the extent of the challenge that sustainability is likely to pose. For the entire sector, or for specific subsectors, programmes or facilities, a good description of sources and uses of funds in the health sector is needed to generate this indicator. Where a substantial role for donors is measured, an assessment of the long term capacity and willingness of the donor to continue to

contribute to the health sector of the country concerned may be needed. Where donor support takes the form of loans--wholly or partly--the contribution to the country's overall debt burden should be considered. At national level, documentation of the contributions of donors should be available, although some donor finance, for example church contributions to mission health facilities, may not be channelled through the Central Government and may be more difficult to ascertain. This information can be supplemented with information gleaned from facility-based studies.

The WHO manual on drug policy indicators (Brudon-Jakobowicz, Rainhorn and Reich 1994) defines some measures that are examples of indicator [b]. To assess trends in government expenditure on pharmaceuticals, they suggest comparing annual government spending on drugs per capita to the inflation-adjusted average of the same measure for the three previous years (p.123). To provide an assessment of the extent to which governments are dependent on international aid to finance or provide pharmaceuticals, they suggest calculating the percent of total government drug expenditures financed from foreign aid, and then comparing that figure to that of previous years to determine if the trend is increasing or decreasing (p.125). While these two indicators examine financial sustainability from the perspective of one key health system input, it is apparent that these calculations could apply to the total of government health spending and international health assistance, not merely pharmaceuticals.

If it is intended that particular facilities become self-financing or increase the extent to which they rely on own-generated revenues (e.g. from user fees and insurance reimbursements), cost recovery ratios [c] are an adequate measure of success in achieving the objective. In comparing cost recovery ratios, it is important to ensure comparable numerators and denominators, and in identifying self-financing as an objective of policy, to identify the intended definition of 'self-financing'--whether, for example, it is intended that salary costs or capital depreciation should be included in the denominator. There is no single 'right' definition of cost recovery ratios or 'self-financing'. The evaluator should choose what he or she wants to measure according to the specific purpose of the analysis. Still, consistent with the definition of sustainability provided above, it is important to recognize that conclusions with respect to changes in the 'sustainability' of a single institution may not be generalizable to the health system as a whole because of the potential for shifting resources across facilities and programmes.