MANAS HEALTH POLICY ANALYSIS PROJECT

Policy research paper number 1

METHODS FOR EVALUATING EFFECTS OF HEALTH REFORMS

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

and

Joseph Kutzin, WHO

November 2000

World Health Organization
Ministry of Health, Department of Reform Coordination and Implementation

1 Togolok Moldo Street, 720405 Bishkek, Kyrgyzstan. Tel. +996 (312) 660438, 666244
Fax +996 (312) 663649, e-mail: moh@manas.elcat.kg

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.

NOTE: This paper is an excerpt and reprint of a paper produced for the WHO *Current Concerns* series. The original paper was printed in 1997 and can be obtained from WHO Headquarters in Geneva with the following reference: *Current Concerns*, ARA paper number 13, WHO/ARA/CC/97.3.

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. 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.

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. Just because an indicator of these objectives 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. 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 Zairian 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 Zairian 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.

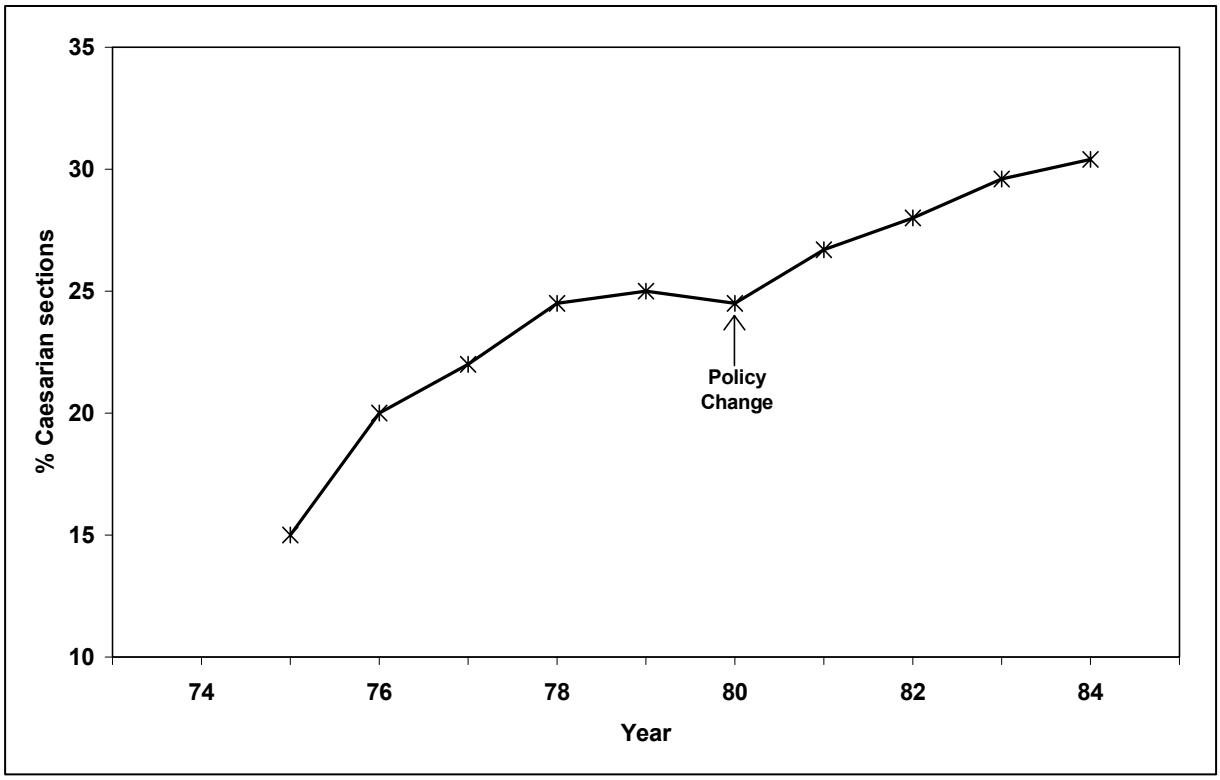


Figure 1. Trends in Caesarean sections – Pelotas, 1975-1984

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.

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

[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.

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).

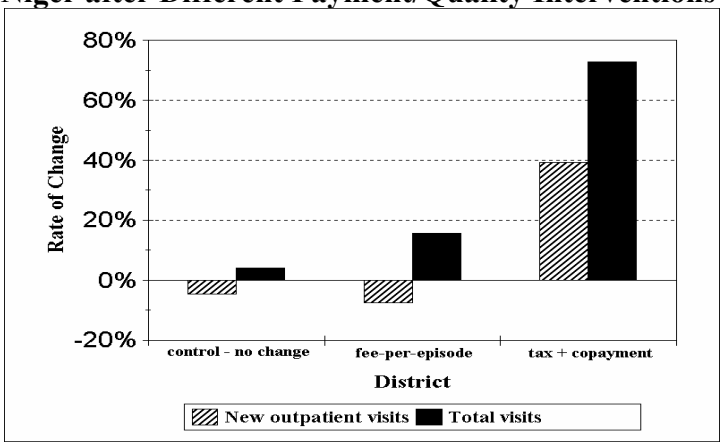
Table 5. Percentage of sick people using a health centre before and after user fee introduction accompanied by quality improvement

	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

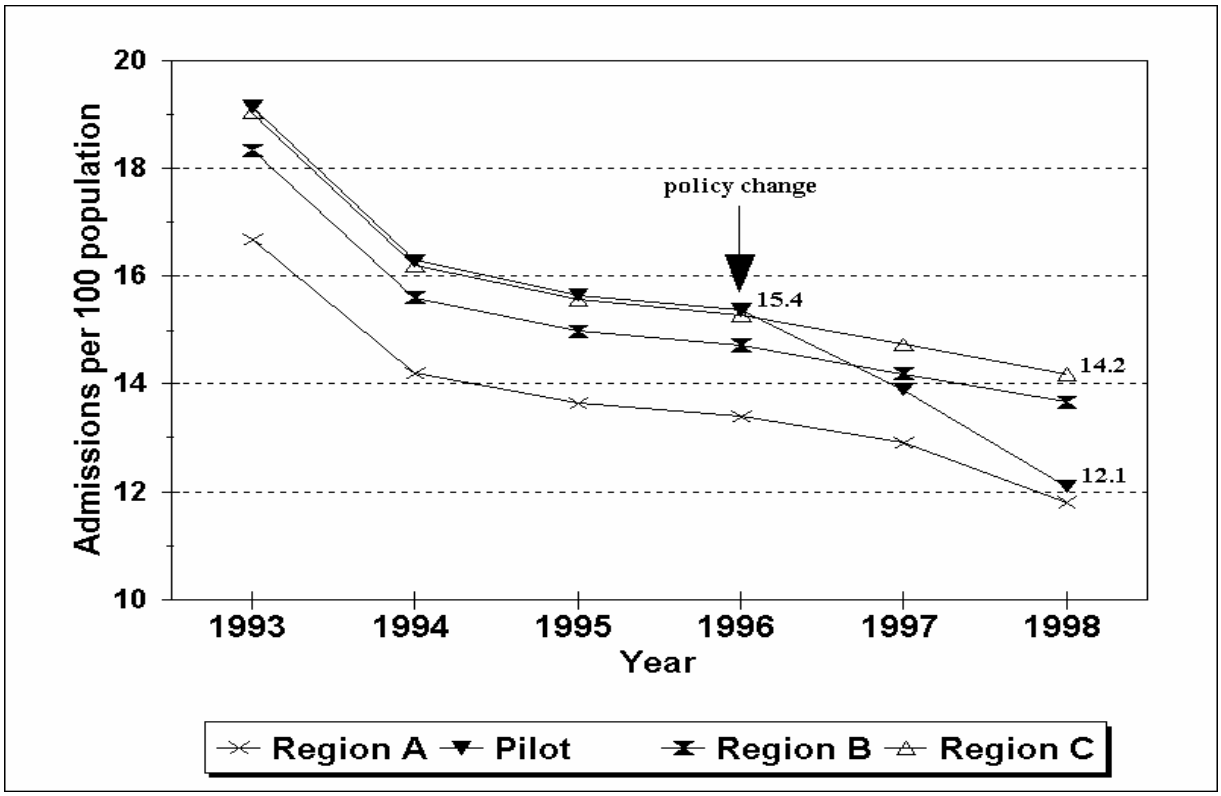
experimental districts, quality and management were enhanced through provision of initial stocks of essential drugs, the introduction of standardized diagnostic and treatment protocols, and the introduction of financial and drug stock management systems. The indicator that they used to measure the utilisation 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 2. Change in Utilization 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 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.

Figure 3. Hospital Admission Rates, Selected Regions in Kyrgyzstan (Hypothetical)

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.

2.3 Generalizing from the findings: some caveats

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.

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.

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.

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.

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

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.

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.

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.

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.

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

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.

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.