

PROPOSED STRATEGIES FOR HEALTH SYSTEMS PERFORMANCE ASSESSMENT

SUMMARY DOCUMENT

Proposed Strategies for Health Systems Performance Assessment	1
Summary Document	1
I. Introduction	2
II. A Framework for Health System Performance Assessment	4
A. Overview	4
B. Framework Proposed in the WHR2000	4
C. Commentaries on the Framework and Proposals for the next round of Performance Assessment	7
III. Inputs	11
A. WHR2000	11
B. Commentaries on the WHR2000	12
C. New analysis	12
D. Proposals	14
IV. Functions	14
A. General Issues	14
B. Financing	15
C. Service Provision	17
D. Resource Generation: Human Resources and Capital	25
E. Stewardship	29
V. Health System Contribution to Social Goals	31
A. Average Level of Population Health	31
B. Health Inequality	39
C. Fairness of Financial Contributions	44
D. Responsiveness	54
E. Composite Goal Attainment	61
VI. Efficiency	64
A. WHR2000	64
B. Commentaries on the WHR2000	65
C. New Analysis	65
D. Proposals - Efficiency	70
VII. Measurement Challenges	70
A. Comparable Estimates Using the Best-Available Evidence	70
4. Proposals	74

B.	Data Collection Strategies and The World Health Survey	75
VIII.	Increasing the Policy Relevance of Health System Performance Assessment	81
A.	Communicating Results to Policy-Makers	81
B.	Using Health System Performance Assessment to Improve Performance	82
IX.	Reference List	87

I. INTRODUCTION

Policy makers have long been concerned with improving the performance of their health systems (1;2) with reforms targeting all system functions - financing, provision, stewardship and resource generation.(1;3-5) An increasing number of studies have assessed the impact of reforms in different settings (e.g.(6;7)), but these studies have used varying frameworks and methods to assess and measure the effect of changes in policies and strategies. This makes it difficult to separate out the true variations in impact from variations stemming from the different methods that were used.

WHO has, therefore, developed a wide ranging work plan to:

- develop the scientific basis to ensure that its technical advice on health system development is based on the best available evidence;
- assist Member States, as requested, to improve their capacity to obtain appropriate evidence, to analyse it, and to use it to improve the performance of their own systems.

One component of this work was to summarise and disseminate the available evidence on the links between health policy, system design and system performance in the World Health Report 2000.(8) Because no single framework had been used consistently to assess and compare performance across settings and over time, the WHR2000 proposed a framework defining the intrinsic goals to which health systems should contribute, and the four key functions of health systems which could be used to improve performance. It showed that the available evidence on exactly how this could be done was limited, inconsistent and inconclusive in many areas.

It argued that the evidence would remain inconclusive unless efforts were made to routinely measure the extent to which the goals were attained and the efficiency with which resources were used to achieve these goals. Without the ability to measure goal attainment in a consistent way across time, it is not possible for countries to monitor and improve the performance of their own systems. Without the ability to measure outcomes in a comparable way across countries, it is not possible to identify which policies and system design options are more likely to be associated with high levels of attainment, and which are not. To demonstrate that it was possible to do this, the Annex tables measured health system contributions to goal attainment and efficiency for all 191 Member States of the Organization.

The uncertainty around these estimates varied according to data availability and quality and uncertainty intervals were reported for all key outcomes. The rationale behind this approach was that policy-makers require timely information to guide their current decision-making. They cannot wait for the scientifically best method to be developed and validated at some point in the future, but require estimates based on the best methods scientifically possible at this time.

The framework and methods used in the WHR2000 were not proposed in isolation, but built on decades of published work on different aspects of health system development. During the course of writing the report and preparing the Annex tables, the input of experts from around

the world, both individually and through technical consultations, was sought. These influenced the development of framework and methods proposed in the Report

The release of the Report engendered considerable interest and debate, from governments, policy-makers and the academic community. Because of this interest, which reflects the perceived importance of the topic, the Director-General undertook to report on the health system performance of Member States regularly, at 2-yearly intervals. Before the next round of performance assessment is to be released in October 2002, there would be a technical consultation process and a small Advisory Group would be appointed to help her monitor WHO's support for the assessment of health system performance. The 107th Meeting of the Executive Board (EB) endorsed these steps in January 2001 and requested WHO to "initiate a scientific peer review of health systems performance methodology as part of the technical consultation process including updating on methodology and new data sources relevant to the performance of health systems".

The terms of reference of the PRG are to:

1. Review the scientific merit of methods proposed by the WHO secretariat for the next round of health system performance assessment, building on the suggestions made in the technical, regional and country consultations, in ongoing research and the general academic debate;
2. Propose revisions, as necessary, to the method that improve their scientific merit, and work with the WHO secretariat to assess the feasibility and impact of any revision;
3. Advise the DG on the scientific merit of the final methods emerging from this process.

Accordingly, this document presents to the PRG a summary of the Secretariat's proposals for the methods to be used in the next round of performance assessment. These proposals draw on the consultations and the associated debate in the scientific literature that have taken place since the publication of the report. Because of the richness of this debate, a considerable number of additional background documents are provided separately.

The background documentation is organized into three parts. This summary document is the first. The second part summarizes the results of the formal consultative process, with reports of the six regional consultations (prepared by the regional offices), (9-14) seven topic-specific technical consultations, (15-21) and a Committee of Experts on Measurement and Classification of Health (22). Part three of the background documentation is the most extensive. It provides copies of published commentaries and new analytical work on the various technical issues relating to the framework, measurement of health system performance and its components, and ways to enhance the performance of health systems using the best available evidence.

The remainder of this summary document is organized around the technical issues and proposals relating to them. Section II presents the suggested framework that guides all subsequent measurement and policy advice. Then proposals related to the inputs to the health system, the functions, and the outcomes follow in sections III to V. The concept and measurement of efficiency is discussed in section VI. Measurement challenges are the focus of section VII, while the final section focuses on ways to increase the policy relevance of health system performance assessment so as to ensure the development of more performant health systems.

II. A FRAMEWORK FOR HEALTH SYSTEM PERFORMANCE ASSESSMENT

A. Overview

Two fundamental principles drive the definition of the health system and the framework for analysing system performance proposed in the WHR2000. The first is that the stewards of the health system should be concerned with promoting actions that improve health, even if direct control for this action lies beyond the narrow mandate of personal curative services, the area which has traditionally been the concern of the medical profession. The second is that a clear focus on outcomes is required if the potential of health systems is to be fully realised. It is also desirable to have information on process indicators showing progress toward reaching final outcomes, but regular evidence on the extent to which the system is contributing to social goals, and the extent to which it is making efficient use of the resources available to it, is an imperative.

B. Framework Proposed in the WHR2000

1. *Definition of the Health System*

The WHR 2000 defined the health system to include all actors, institutions and resources that undertake **health actions** – i.e. all actions whose **primary intent** is to improve health. (8;23) It is a broader definition than the health actions typically under the direct control of a Ministry of Health, and encourages the stewards of the health system to focus on the delivery of key personal and non-personal health services, as well as to be effective advocates for inter-sectoral activity on a range of actions aimed specifically at improving health. These would include seat belt legislation to reduce road traffic accidents and taxation to reduce the consumption of tobacco products.

2. *Goals*

To be outcome-focused, it is necessary to define a set of goals and to measure progress toward achieving them. The health system contributes towards many outcomes that are socially desirable, including improving health, educational attainment, and individual incomes. It is necessary to develop a set of criteria to determine which goals are **intrinsically valued** and which of those should be measured routinely.

Two criteria were established to define a goal as **intrinsically valued**.

- It is possible to raise the level of attainment of the goal while holding the level of all other intrinsic goals constant – i.e. an intrinsic goal must be at least partially independent of all others. Partial independence does not mean completely independent, only that it is possible to imagine independent variance in the goal compared with other intrinsic goals.
- Raising the level of attainment of an intrinsic goal is always desirable – e.g. more is always better than less. If the levels of attainment of other intrinsic goals are kept constant and raising the level of attainment of a given goal is not necessarily desirable, it is probably an instrumental and not intrinsic goal.

An **instrumental goal** is something that is desirable because it contributes to attainment of an intrinsic goal. More is not necessarily better than less, holding attainment of the intrinsic goals constant.

To warrant measuring attainment of an intrinsic goal regularly, two additional criteria were proposed. The health system must be able to make a large enough contribution to the goal to

warrant the expense of measuring it regularly and it must be feasible to measure the health system impact on a regular basis.

Using these criteria, three intrinsic goals were identified. The **defining goal** of a health system is to improve **health**, where there are two components. The system should seek to improve the average **level of population health** as well as trying to reduce **inequalities in population health**.

The second intrinsic goal was to enhance the **responsiveness** of the health system to the legitimate expectations of the population for the non-health improving dimensions of their interaction with the health system. Responsiveness does not include expectations for the health improving dimensions of their interactions which are fully reflected in the first goal. "Legitimate" was used to recognise that some individuals may have frivolous expectations for the health system which should not be part of the articulation of this goal. Responsiveness has two key sub-components: respect of persons and client orientation. Again, we are concerned not just with the **average level of responsiveness**, but also with **inequalities in responsiveness** within the population.

The third intrinsic goal was the **fairness in financial contributions** to the health system. To be fair, households contributions to finance the health system should represent an equal sacrifice. Equal sacrifice would mean that no household would become impoverished or pay an excessive share of their income to finance the health system. Equal sacrifice also means that the poor households should contribute a smaller share of their income than rich households.

The WHR2000, therefore, measured 5 components of the 3 intrinsic goals (Table 1). Each intrinsic goal meets the criteria established above. More health (more responsiveness and more fairness of financial contributions) is always better than less holding attainment on the other goals constant. Even though improved responsiveness can encourage people to seek care, thereby improving health, it is possible to improve responsiveness by providing improved amenities, for example, with no impact on health outcomes. The three goals are partially separable, and the impact of the system is sufficiently large to warrant measuring them regularly. Although it was recognised educational attainment and income-earning potential might meet the criteria of an intrinsic goal, it was judged that it was not practical to try to routinely measure and report the contribution of the health system to those goals.¹

Table 1. Intrinsic Goals to which the Health System Contributes

	LEVEL	DISTRIBUTION
HEALTH	√	√
RESPONSIVENESS	√	√
FINACIAL CONTRIBUTION		√
	QUALITY	EQUITY

A number of instrumental goals, such as the coverage of health services, were discussed in the text of the WHR2000. But attainment on these goals was not measured in the annex tables.

¹ The Commission on Macroeconomics and Health, established by the Director-General to advise her on the contribution of health and economic development, will release its report in December 2001. One of its main findings is that there is convincing macroeconomic and microeconomic evidence that investing in health will increase economic growth.

3. *Health System Functions*

The WHR2000 defined four basic functions which contribute to determining observed levels of goal attainment – financing, service provision, resource generation, and stewardship.

Health system financing is the process by which revenues are collected, accumulated in fund pools and allocated to specific health actions. It can be subdivided into three sub-functions: revenue collection, fund pooling and purchasing. **Service provision** refers to the way inputs are combined to allow the delivery of a series of interventions or health actions. It includes personal and non-personal health services. The former are services consumed directly by an individual, preventive, diagnostic, therapeutic, or rehabilitative. The latter are actions that are applied either to collectivities (e.g., mass health education, legislation) or to the non-human components of the environment (e.g., basic sanitation).

Health systems are not limited to the institutions that finance or provide services, but include organisations that produce inputs to those services, particularly human resources, physical resources such as facilities and equipment, and knowledge. Included are educational institutions, research centres, construction firms, and the vast array of organisations producing technologies such as pharmaceutical products, devices and equipment. Strategies for **resource generation**, the third function, can be critical to the performance of the system.

Stewardship is a neglected function in most health systems, extending beyond the conventional notion of regulation. It involves setting, implementing and monitoring the rules of the game for the health system; assuring a level playing field among all actors in the system (particularly purchasers, providers and patients); and defining strategic directions for the health system as a whole.

The WHR2000 summarised, in the text, the available evidence relating these functions to goal attainment and efficiency. However, it did not define indicators of how well the functions were being undertaken and no attempt was made to measure process indicators linked to the functions.

4. *Inputs*

The ability to produce desirable outcomes depends on the inputs available to the system and whether scarce resources are used as efficiently as possible. So one reason for measuring inputs is to focus on whether systems are achieving as much as they could for the available resources. The second reason is to help understand why they might not be achieving as much as they could. One possible explanation is that they are using an inappropriate mix of inputs – personnel, equipment, pharmaceuticals etc – for their systems.

In the WHR2000, the function of resource generation was discussed in the text, focusing mainly on human resources. The only input reported was health expenditure per capita, broken down by public and private sector contributions. This was reported in US dollars at official exchange rates and in international dollars converted at purchasing power parities.

5. *Performance (Efficiency)*

The final concept proposed in the framework was called performance, equivalent to the concept of efficiency. It is a function of the system's contributions to the intrinsic goals taking into account the inputs used to achieve them. The vertical axis in Figure 1 measures goal attainment while the horizontal axis is the quantity of inputs used by the system. M is the maximum achievable level of outcome for the inputs, taking into account non-health system determinants. L is the minimum possible level of goal attainment in the absence of inputs,

necessary because health would not be zero in the absence of the system.² Performance was defined as the ratio of attainment (above the minimum) to the maximum possible attainment (also above the minimum) – i.e. what proportion of the potential health system contribution to goal attainment is actually achieved for the observed level of resources. For a country at A it is $e/(e+f)$. Full details of the methods are defined in an associated background paper, and a summary is included in section VI.

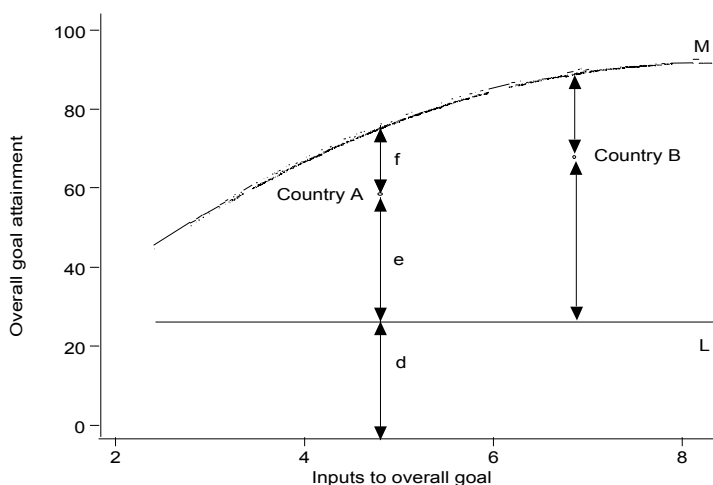


Figure 1: Defining Health System Efficiency

C. Commentaries on the Framework and Proposals for the next round of Performance Assessment

Debates and suggestions relating to the way that goal attainment should be measured, and the way efficiency should be defined and measured, are described in subsequent sections of this document. Here, philosophical issues relating to the framework are described and proposals are made to the PRG for the next round of performance assessment. **(24) The proposals are in bold type.**

1. Definition of the Health System

A number of alternative concepts of the health system have been suggested subsequent to the release of the WHR2000. The narrowest draws the boundaries tightly around the activities under the direct control of the Ministry of Health. In some countries this includes largely personal medical services and excludes activities such as the marketing of insecticide impregnated mosquito nets or taxes designed to reduce the use of tobacco or alcohol products. It is depicted as the smallest circle in Figure 2. This definition encourages the stewards of the health system to focus only on a very narrow set of personal curative services. The second definition is more inclusive, corresponding to the second smallest circle in Figure 2. In it, the system is defined to include personal medical and non-personal health services but not inter-sectoral actions designed to improve health. The type of inter-sectoral actions in which WHO has long been engaged, such as safe water and sanitation programmes, would not be included. Again, stewards of the health system would have no incentive to advocate for the introduction of anti-smoking legislation, for example, if they took responsibility only for this set of health actions.

² Inputs, particularly health expenditure, are likely to be positively correlated with most non-health system determinants of goal attainment such as education and housing. So the minimum possible across countries would be seen to rise with increases in inputs due to this correlation.

The third definition corresponds to the WHR2000 definition. It incorporates any action whose primary intent is to improve health. This is broader than personal medical and non-personal health services. It includes inter-sectoral actions which the stewards of the health system could take to advocate for health improvements in areas outside their direct control – e.g. legislation to reduce fatalities from traffic accidents. This is shown by the largest circle in Figure 2.

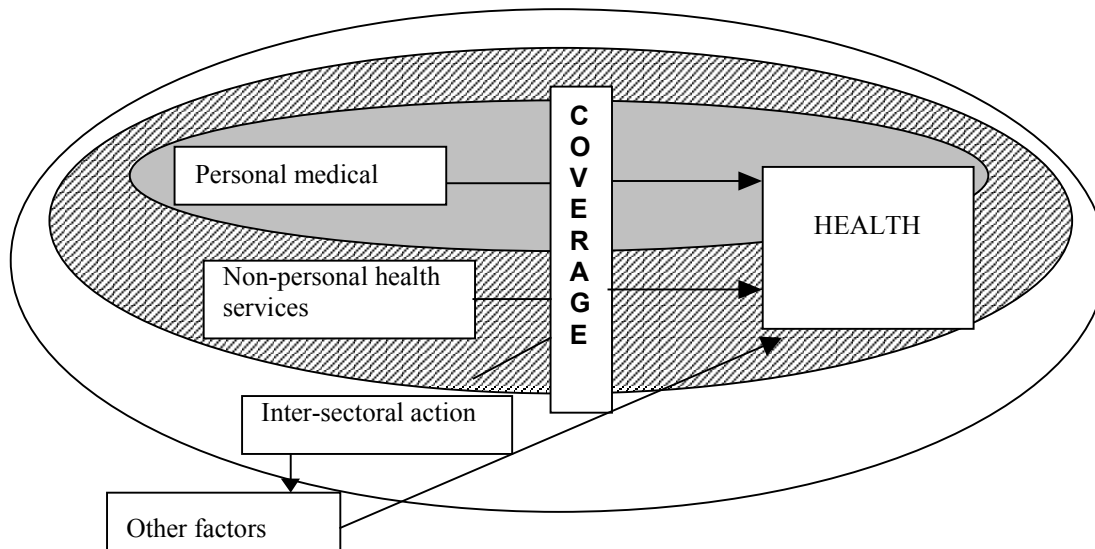


Figure 2: Boundaries of the Health System

The final option is to define the health system to include all actions that might contribute to improving health. This includes every box in Figure 2 because all areas of human activity – e.g. education, industry, tourism and agriculture – can influence health. There is no longer any operational distinction between the health system and any other system and it would not be possible to assess the performance of the health system.

Proposal:

WHO proposes that the WHR2000 definition should be retained because it encourages stewards of the health system to focus on a definable set of actions whose primary intent is to improve health.

2. Goals

Since the publication of the WHR2000, there has been general consensus that health systems should be concerned with health, that they should be responsive to people’s non-health related expectations, and that financial contributions should be distributed fairly across households. There has been considerable debate about what is the appropriate indicator to use to measure these goals, and on the data sources. This has resulted in considerable refinement and elaboration of the methods and improvements in the data sources which are described in section V.

There has also been debate about whether other intrinsic goals should be included for routine monitoring. Indeed, two opposing views were expressed during the consultation process. One was that health systems should be concerned with, and judged on, contributions in terms of health alone. Proponents of this view argued that responsiveness and fairness of financial contributions were desirable characteristics but they were not goals on which system performance should be judged. The opposing view was that additional intrinsic goals should be measured in addition to health, responsiveness and fairness in financial contribution including the impact of the system on education and income. In addition, there was debate about whether coverage, access and the level of financing were instrumental or intrinsic goals.

Proposals:

WHO proposes that coverage, access and participation are instrumental goals. More is only better than less to the extent that more contributes to one of the defined intrinsic goals. More access would not be desirable if it did not improve health or responsiveness, for example. It would in fact be a waste of resources if increasing access had no impact on health or responsiveness.

This being said, many of the regional consultations suggested that policy-makers need information on certain instrumental goals so that they can monitor the extent to which their systems were likely to achieve the desired intrinsic goals. WHO proposes to the PRG that this suggestion needs to be incorporated into routine health system performance assessment and ways in which this might be done are described in section IV.

The impact of the system on education and income do meet the two criteria for an intrinsic goal, and the links between poor health and poverty are well established. However, we argue because of the measurement challenges involved that it would be more appropriate to measure the impact of the health system on educational attainment and income as part of periodic research studies rather than regular performance assessment. In conclusion, we propose that the three goals and five components of Table 1 be retained.

The level of finance is more difficult. It is theoretically possible to define an appropriate level for each country – the point at which the marginal benefit of increasing expenditure on health just equals the marginal benefits foregone from switching resources from non-health spending. Then, moving closer to the ideal level would be an intrinsic goal. However, defining this ideal level in practice has never been done. For the moment, WHO proposes to continue its work aimed at defining the appropriate level of expenditure, but that it would not yet be incorporated into routine health system performance assessment as an intrinsic goal.

3. Attribution and Measurement

Many of the consultations pointed out that goal attainment, particularly in health, is not simply a function of health actions. It is also influenced by non-health system actions such as poverty alleviation, education and agricultural development policies. In Figure 2, the box labeled “other” captures these activities.

Multivariate statistical analysis allows the concept of outcome measurement to be separated clearly from the assessment of causal attribution. Overall child mortality and changes in it, for example, can be measured and then the causes explored using multivariate statistical analysis. This allows all possible hypotheses of causation to be tested.

An alternative approach implicit in many of the comments received since the publication of the WHR2000 is to define partial indicators of overall goal attainment which are believed to be uniquely or largely determined by the activities of the Ministry of Health. In this approach, child mortality due to vaccine preventable diseases would be measured in preference to overall child mortality because the Ministry of Health can control the coverage of routine vaccination programmes but cannot control the fact that some children die from malnutrition associated with poverty. In a similar way, one might argue to measure mortality due to medical errors instead of adult mortality, because overall adult mortality may be determined by factors outside the control of the health system.

This approach focuses attention only on determinants that are identified *a priori*, rather than allowing analysts to explore all possible causes. We agree that some of these indicators might be useful to measure in different settings, but not at the expense of measuring the system’s contribution to goal attainment.

Proposals:

WHO proposes that:

- 1. It is important to measure the system's contribution to the desired final outcome. This gives the power to explore all possible hypotheses of causation. For example, inequalities in health outcomes would be measured so that contribution of different factors such as income, social class and interventions provided by the health system could be analyzed.**
- 2. Regular measurement of a limited set of process indicators or intermediate goals should also be undertaken, in line with the recommendations of several regional and technical consultations.**

4. *Inputs*

During the consultative process, no one seriously questioned the use of health expenditure per capita as an aggregate indicator of the inputs available to the system. Since the report was published, many countries have provided updated estimates of health expenditure and richer detail of its breakdown by category, and this is outlined in section III.

Some concerns were raised about the source of purchasing power parity exchange rates used to convert local currency units into international dollars, but these are technical questions discussed in the background documents. More important has been the argument that it is not just health expenditure in the current time period that is important in determining outcomes, but past health expenditure, as well as the stock of capital available at the beginning of the period (also determined by past expenditures).

Another concern since the release of the report, is that while health expenditure is critical, additional information on the types of inputs available to the system and how they are used would be useful in identifying possible reasons for poor performance.

Proposal:

Proposals related to the timing of inputs and outcomes are taken up in section VI. WHO proposes to explore the feasibility of estimating the quantities of different types of labour inputs (e.g. nurses, doctors) and the stock of capital (e.g. health facilities, equipment) available to the health systems of our 191 Member States. This is discussed in section III.

5. *Health System Functions*

As stated in earlier sections, it was almost universally agreed at the consultations that while information on intrinsic goal attainment was important, it was only a starting point. It provides information on how well the system is performing, but no indication on how to improve performance. The request was for indicators that allowed policy-makers to "drill down", to discover possible causes of poor performance and ways in which it might be improved. The other simile that was used was that of a car dashboard.⁽⁹⁾ The ultimate goal of stepping into a motor vehicle is to arrive at a destination, but the indicators on the dashboard can be useful in indicating to the driver how long it is likely to take, and whether there are major problems which might prevent goal attainment.

Proposal:

WHO proposes to measure routinely a set of instrumental goals linked to each of the four functions as well as selected attributes of these functions (described in section IV). The purpose of the latter is to identify relationships between the way the functions are organized and undertaken and health system performance.

6. Performance and Efficiency - Terminology³

In the WHR2000, the term “performance” was used as a synonym for “efficiency”. This proved to be confusing and a number of the regional consultations recommended that “health system performance assessment” should be defined to include the range of activities from measuring goal attainment, the efficiency of input use, and the way the system is functioning. The term “efficiency” would then be used in the narrower sense of how well resources are used to produce the desired outcomes.

Proposal:

WHO proposes that “performance” should be defined to include the range of activities from measuring the system’s contribution to goal attainment, the efficiency of input use, and the way the system is functioning. The term “efficiency” should be used to define how well resources are used and combined to produce the desired outcomes.

III. INPUTS

The previous section described the framework proposed for the analysis of health system performance. This section is the first of a number of sections outlining ways of measuring health system inputs and the system’s contribution to goal attainment.

A. WHR2000

Information on health expenditure is important in two ways. Firstly, the current availability of resources constrains the extent to which outcomes can be improved. Secondly, some countries seem to be able to achieve more for any observed level of expenditure, and understanding why this is the case is critical to efforts to improve health system performance. That type of analysis requires information on outcomes and expenditures at the same time.

For the WHR2000, less than 70 countries had produced relatively complete National Health Accounts (NHA), and less than a third of them traced expenditures through the main components of health care financing: resource mobilization; resource allocation/purchasing; and provision of services. For the other 121 countries, WHO had to seek information on health expenditures from whatever sources were available, a monumental effort. This involved interaction with national authorities as well as searching for data from consultant’s reports, researchers, and health expenditure surveys. Estimates of health expenditures were reported for all 191 Member States, disaggregated by source – public (tax funded and social insurance) versus private (out-of-pocket and total).

In the text of the WHR2000, the function of resource generation and the ways it could influence performance were described. However, no attempt was made to disaggregate health expenditure by type of input – labour, capital, consumables, for example – or to report the quantities of these inputs available to the different health systems.

³ Technical commentaries relating to the definition and measurement of efficiency are described in section 6.

B. Commentaries on the WHR2000

Some of the estimates of health expenditure reported in the WHR2000 were criticised on the grounds that they were based on outdated data, and as stated earlier, some countries disagreed with the estimates of GDP, exchange rates (official or PPP), or population used. Others questioned the categorization of specific type of expenditure, for example whether social insurance is public or private, and suggested that it would be useful to have information on the quantities of different types of inputs available to each country's health system. Finally, a number of countries requested technical assistance to undertake full or partial National Health Account estimations.

C. New analysis

In response to these suggestions, there has been continual interaction with national authorities and other international agencies to update estimates using the most recently available data. The resulting estimates for 1998, and revisions for 1997, were sent to countries for comment and changes agreed with them were reported in the annex tables to the WHR2001.(25;26) A new breakdown of expenditure was also reported – external resources for health as a percentage of total public expenditure. In 1998 the world spent I\$43,1 trillion on health.(27) Not only is the level of health expenditure per capita lower in poorer than richer countries but the proportion of GDP devoted to health is also lower (Table 2). Expenditures per capita for the countries grouped by income varied from just over I\$24 to just over \$2,000. At the country level the variation is even wider than shown in Table 2 – country-level expenditure varied from \$11 to \$4,055 per capita in 1998 (I\$).

Income Group (GDP p.c.)	Total Health Expenditure		Per Capita Health Expenditure		Share of GDP (PPP) (%)
	(US\$)	(I\$)	(US\$)	(I\$)	
<1000	5,414.78	9,985.18	13.11	24.17	3.38%
1000-2200	40,824.02	156,437.60	22.87	87.65	4.52%
2200-7000	205,256.70	518,709.80	81.70	206.46	5.09%
>7000	2,388,315.00	2,387,353.00	2,043.02	2,042.20	9.63%
Total	2,639,810.00	3,072,486.00	448.99	522.58	7.94%

Table 2. Health Expenditures 1998

There is also considerable variation in the share of total health expenditure provided by different sources (Figure 3). Out-of-pocket payments account for over 70% of total health expenditures in some parts of the South-East Asian Region of WHO where private health insurance is almost non-existent⁵. On the other hand, private health insurance accounts for over 20% of all health expenditure in the part of the American Region which includes the USA.

⁴ I = an international dollar is a currency unit which has the same purchasing power in a given economy as US\$1 will have in the US economy.

⁵ WHO regions are described in Annex table 1. In Figure xxx the components are external (ext), private insurance (pvt), social insurance (ssh), tax based payments (tax), and out-of-pocket payments (Oops).

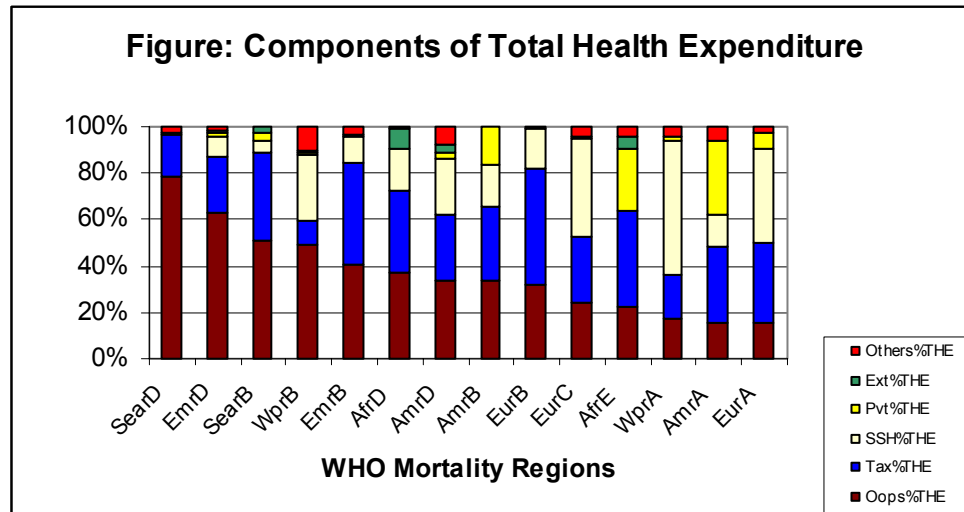
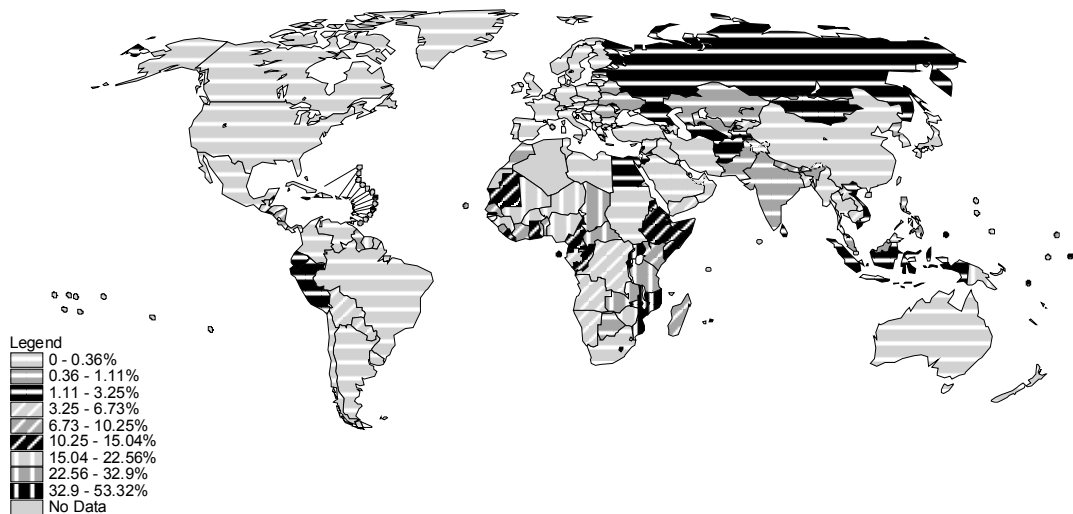


Figure 3 : Components of Total Health Expenditure

External funding for health also plays an important role in some regions, and the variation shown in Figure 3 is even more obvious at country level. Figure 4 depicts external funding as a percentage of total health expenditure by country. In some countries, external resources account for over 50% of total health expenditure.

Externally funded health care, 1998 (measured share in total health spending)



The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement. WHO/2001. All rights reserved.

Evidence and Information for Policy



On the process side, ongoing talks with the OECD have resulted in estimates for OECD countries being harmonized between the two organizations. In addition, WHO and OECD have joined with the World Bank, the US Agency for International Development and the Swedish International Development Co-operation Agency to develop a Producer's Guide, a step-by-step guide for countries interested in developing their own NHA estimates.

D. Proposals

WHO proposes to:

1. Provide additional breakdowns of expenditures to the extent possible. It is probably feasible to provide breakdowns of expenditure by type of care (inpatient, outpatient, long term care), by inputs (capital, labour etc), and by provider. In the longer run it should be possible to breakdown expenditure by beneficiary group (by sex and income) and possibly by type of health problem.
2. Develop ways of estimating the numbers and types of health personnel available in each country, both public and private sectors.
3. Develop and apply methods to measure the physical capital stock available to the health system - the value of the buildings and equipment available to contribute to attainment of the intrinsic goals (see section IV).

IV. FUNCTIONS

The way the health systems translates inputs into outcomes is influenced by the way the functions of the health system are organized and undertaken. This is the focus of section IV.

A. General Issues

The World Health Report 2000 described the four major functions of the health system and summarised the available evidence on the how different ways of undertaking these functions influence health system performance. This evidence was mixed, which is one of the reasons why it was decided to try to measure outcomes in a more consistent way than had been undertaken in the past. Stewardship was considered to be the most important function because it influences the way the other functions are undertaken and makes possible the attainment of the intrinsic goals. But no attempt was made to measure how these functions were organized and performed in the different Member States, nor the extent to which they contributed to attainment on the intrinsic goals. One of the strongest recommendations to emerge from the consultative process was that WHO should develop and test a set of indicators linked to the functions as a way of increasing the policy relevance of performance assessment.

General Proposals:

A parsimonious set of indicators of the performance of functions is presented in this section. The set should be small enough so that the health information systems of poor countries can routinely measure and report them. Countries with greater resources might choose to report more.

WHO proposes to include two types of indicators:

- i. indicators of instrumental goals related to the functions - e.g. where the correlation between attainment of that indicator and attainment of an intrinsic goal is well established;
- ii. indicators of the organization or performance of the function where there is an hypothesized, but not yet proven, relationship with intrinsic goal attainment.

Regular measurement of the second type of indicator across settings and over time will help to build the evidence-base for the development of health system policy advice.

Specifically we ask the PRG to consider if WHO should:

- Develop measurement strategies for the indicators proposed below;

- **Report on those where it is feasible and affordable to measure indicators that are reliable, valid and comparable;**
- **Continue work to establish the relationship between the indicators and health system performance.**

The rest of this section will, therefore, describe the type of indicators that might be measured for each of the functions.

B. Financing

A good health financing system contributes to goal attainment by generating enough resources to adequately support the health system at the desired level of output. (28) It can also be used to modify the incentives faced by those who pay and those who receive funds for either the provision or utilization of health interventions. Financing systems can, at least in theory, affect health outcomes – people who cannot pay for services in fee-for-service systems would have lower health outcomes than in other systems. They can clearly influence the fairness of financial contribution goal by making catastrophic payments more, or less, likely. Nevertheless, there is substantially less known about the direct or indirect impact of the financing system on health outcomes and responsiveness.

Public policy on the financing function needs to address a range of trade-offs related to mobilizing funds, utilization of services, provider incentives to produce good quality services, pricing policy in relation to costs, and distribution of the financial burden. Current evidence suggests that the impact of any policy introducing changes in one or more components of the financing function will be affected by the socioeconomic context, including the national income level, performance of other sectors (e.g. road safety, legal enforcement of contracts), public institutional capacity, professional standards, and general levels of education.

Subsequent to the publication of the WHR2000, WHO has responded to the demands from countries for more explicit statements about the way that financing policies affect health system performance by beginning to develop a WHO Health Financing Policy. This work is not yet finished, but it is taking its evidence from existing studies as well as the analysis of new data. It is focusing on policy issues that are relevant to particular clusters of countries. For example, questions regarding the appropriate level of health spending in high-income countries are generally more concerned with cost-containment; whereas in the lowest income countries, the chief concern is with ways to mobilize additional resources to improve health.

One possible list of dimensions and potential indicators is found in Table 3. Many of the proposed indicators can be obtained or calculated from data that already exist or are collected by other organizations. For example, public sector spending as a share of GDP is reported by the International Monetary Fund. Out-of-pocket payments for health uses data collected in health expenditure surveys. Others will only be available if new surveys are undertaken, particularly those aimed at collecting information from service providers, and the World Health Survey, described in section VII, is compiling a module which can be used to measure some of these indicators and explore their validity, reliability and comparability.

Special mention should be made of the work to define an appropriate level of health expenditure for inclusion in the WHO Financing Policy. In theory this is simple to do – it is the point at which the marginal benefit of investing more in the health system just equals the marginal benefit of investing in any other system (e.g. education, defense etc.). However, we have not yet found a way to operationalise this for our Member States and the input of the PRG on this difficult issue would be welcome.

Proposals - Financing

WHO should explore the validity, reliability and comparability of a small set of suggested measures, and define a small set of indicators which can be measured routinely. The indicators suggested in Table 3 are a starting point. We invite the PRG to comment on which ones they feel are the most important.

a) Indicators	Purpose
b) Revenue Collection	
<ul style="list-style-type: none"> • The formal sector share of GDP • Natural resource revenues as a share of total public sector income 	Potential resources available to finance public health spending
<ul style="list-style-type: none"> • Public sector expenditures as a share of GDP • External health sector aid as a share of total public health expenditures 	To measure resources specifically available to the public sector
<ul style="list-style-type: none"> • The share of public health expenditures in total public expenditure • Total health expenditure (per capita level and share of GDP) 	To measure public sector allocation decisions, additional resources, and potential constraints
<ul style="list-style-type: none"> • The share of total health expenditures that are prepaid (as against those which are paid out-of-pocket at time of service) 	A broad measure of financial protection against out-of-pocket expenses
c) Pooling	
<p>Means and concentration indices of:</p> <ul style="list-style-type: none"> • Share of copayments to total health expenditure in each pool • Membership in each pool • Per capita spending in each pool 	Measures of the scale, depth of financial coverage, and existence of compensatory mechanisms across pools
d) Purchasing	
<ul style="list-style-type: none"> • Number of purchasers • Mean and distribution of total expenditure across purchasers • Mean and distribution of the number of providers who are contracted or hired by each purchaser 	To characterize the structure of interactions between purchasers and providers
<ul style="list-style-type: none"> • Share of total funds allocated by inputs (e.g. salaries and traditional budgets), outputs (e.g. fee for service) and outcomes (e.g. capitation). 	To measure the financial incentives embedded in payments to providers

Table 3 : Selected Proposals for Measuring and Monitoring the Financing Function

C. Service Provision

Provision refers to the process of combining inputs to produce a set of health actions.(29) The inputs are human resources, physical capital, and consumables. The outputs are personal and non-personal health services and intersectoral health actions. This covers all parts of the health system – government and non-government.

Three areas are critical for the assessment of health service provision: (a) the coverage of critical interventions achieved by the health system, (b) health system inputs, and (c) organizational structure and processes. The outcomes of the health service delivery process will be captured by the measurement of the overall level and the distribution of health and responsiveness.

1. Coverage of critical interventions

The most obvious established indicator of how well the provision function is contributing to the intrinsic goals is the coverage of critical health interventions.(30;31) The link between coverage with critical interventions and levels of population health is so well established that effective coverage should be considered as an instrumental goal of health systems.

We define effective coverage in a particular health system as the total realized health gain from these critical interventions as a fraction of the potential health gain possible if providers performed those interventions at the optimal level for that health system. This applies to personal and non-personal health interventions. In this section, we develop more carefully the definition of coverage and its decomposition into key factors such as availability, accessibility, affordability and acceptability.

For an individual i , coverage C_{ij} can be defined as the probability of receiving a health intervention j conditional on needing health intervention j . For intervention j it is:

$$C_{ij} = F(B_{jk}, Y_i, K_{ijk}, Z_{jik}, R_{jk})$$

This shows C_{ij} as a function of the price of the intervention j offered by provider k to the individual i (B_{jk}), individual's disposable income (Y_i), cultural acceptability of intervention j offered by provider k to individual i (Z_{jik}), physical access of individual i to provider k for intervention j (K_{ijk}), and availability of the required technology to provider k to deliver intervention j (R_{jk}).

It is important to note that we have defined the expression for coverage at the individual level with an intervention as a conditional probability. Such an expression allows us to more precisely define each of the associated constructs that are relevant to the assessment of coverage. It also allows us to define inequality in coverage (see below).

Five types of coverage can be defined based on the conceptual framework described in the background documents and similar models found in the literature:

- Availability coverage;
- Accessibility coverage;
- Affordability coverage;
- Acceptability coverage;
- Effective coverage.

Each of these is a function of various individual and health system factors. Overall effective coverage is the ultimate domain of coverage which links the concepts of coverage with effectiveness of health interventions.

It is useful to be able to assess the limitations imposed on effective coverage by the availability of resources, the physical accessibility of health service providers, the affordability of interventions,

and the cultural acceptability of providers and interventions. Each of these factors can be considered as determinants of effective coverage.

Availability coverage for individual i for intervention j , availability coverage can be defined as the probability of an individual receiving an intervention if the only limiting factor for receiving the intervention was the availability of the technology to deliver the intervention at health providers.

Availability coverage for intervention j for individual i can be represented as:

$$Av_{ij} = 1 - \prod_k (1 - C_{ijk} | (B_{jk} = 0, Z_{jik} \geq \theta, K_{ijk} \leq \tau)).$$

The expression says that availability coverage is a function of the conditional probability for an individual i to receive intervention j from provider k , taking into account the independent probabilities receiving the intervention from each possible provider. The conditional statement says that it is the probability of coverage given a price of zero, cultural acceptability that is greater than some standard and physical distance less than some standard. In other words, it is the coverage that would apply if **(a)** the price of intervention j offered by provider k is zero ($B_{jk}=0$) (so income is not a constraint either); **(b)** the cultural acceptability of intervention j delivered by provider k to individual i is equal or greater than a certain threshold value θ of the latent variable of 'cultural acceptability' ($Z_{jik} \geq \theta$); and **(c)** physical access (measured by distance or time) for individual i to the provider k who delivers intervention j is equal or less than a certain threshold value τ ($K_{ijk} \leq \tau$). The only constraint is availability of the intervention.

Defined at the individual level, availability coverage could be aggregated across interventions to get an overall assessment of availability coverage for an individual:

$$Av_i = \frac{\sum_j Av_{ij} d_{ij}}{\sum_j d_{ij}}$$

In this expression, the availability coverage conditional on needing intervention j is aggregated over interventions taking into account the probability, d_{ij} , that an individual will need the intervention j . An overall health system availability coverage can also be defined by aggregating the availability coverage across individuals:

$$Av = \frac{\sum_j \sum_i Av_{ij} d_{ij}}{\sum_j \sum_i d_{ij}}$$

Accessibility coverage for individual i for intervention j is defined similarly, the only constraint being physical access:

$$Acs_{ij} = 1 - \prod_k (1 - C_{ijk} | (B_{jk} = 0, Z_{jik} \geq \theta, R_{jk} = 1)).$$

It is the probability of coverage with intervention j given that there are no constraints in terms of affordability, cultural acceptability and the availability of the required technology. Accessibility coverage at the level of the individual or for the whole health system can be defined in an analogous fashion to availability coverage.

Affordability coverage for individual i removes any the constraints on the availability of intervention j , access, and cultural acceptability. The only determinants of use are then price and income:

$$Aff_{ij} = 1 - \prod_k (1 - C_{ijk} \mid (K_{ijk} \leq \tau, Z_{jik} \geq \theta, R_{jk} = 1)).$$

Affordability coverage is the probability of coverage for individual i , given that there are no constraints in terms of the cultural acceptability, physical access, and availability of technology.

Acceptability coverage can be represented by the following notation:

$$Aff_{ij} = 1 - \prod_k (1 - C_{ijk} \mid ((B_{jk} = 0, K_{ijk} \leq \tau, R_{jk} = 1))$$

In other words, acceptability coverage for individual i is the probability of coverage with intervention j given there are no constraints other than cultural acceptability.

Each of these constructs is a type of counterfactual statement about what coverage would be if only one factor was the constraint. Coverage itself, however, can be modified by introducing the notion of effectiveness. Effectiveness of an intervention can most easily be thought of as the fraction of potential health gain from an intervention realized for an individual given the actual behaviour of the provider. Taking the notion of effectiveness together with coverage, we can define effective coverage for an individual i with intervention j as:

$$EC_{ij} = \frac{HG_{ij} C_{ij}}{(HG_{ij} \mid P_k = P_{opt})}$$

The numerator is the expected health gain from intervention j for individual i - HG_{ij} is health gain from intervention j ; C_{ij} is the probability of receiving effective intervention j . The denominator is the maximum possible health gain from the delivery of intervention j for individual i if the provider behaves in an optimal manner - P_k describes provider characteristics or performance (technical quality); and P_{opt} is the optimal provider performance possible for a given health system.

We can likewise define effective coverage for an individual across a set of interventions:

$$EC_i = \frac{\sum_j HG_{ij} C_{ij} d_{ij}}{\sum_j (HG_{ij} \mid P_k = P_{opt}) d_{ij}}$$

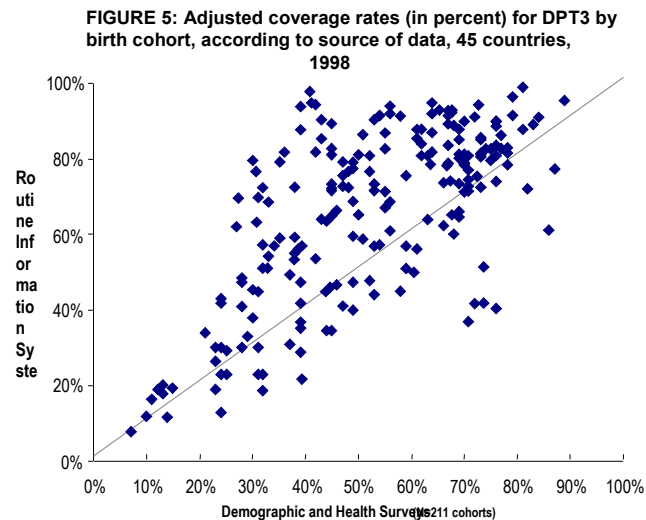
d_{ij} is the probability that individual i has a specific health problem that requires intervention j . The probability of needing the interventions d_{ij} is essentially a weighting factor which gives more importance to interventions against more likely health conditions.

2. Challenges of Measuring Effective Coverage

In theory at least, it is possible to measure effective coverage and decompose it into its components. We propose to explore methods and tools to undertake this but it poses important challenges. Coverage has only rarely been measured for many interventions, and where it has been, the most important source of data has been service registries kept in health facilities.

Some of the well known limitations of data from service registries include:

- Data are incomplete covering mainly public sector activities;
- Data are biased capturing only people who have accessed health facilities;
- Reporting can be haphazard or biased because of a lack of incentives for staff to register cases appropriately;
- Cross-population comparisons, both within and across countries, are difficult due to variations in medical practices and registration procedures.



Despite these limitations, these data are often all that is available. Some programmes, particularly the Extended Programme of Immunization, have invested significant resources in improving the quality of reporting so that they can obtain reliable coverage estimates. However, there are many challenges ahead as the quality of service data remain poor. Figure 5 demonstrates the difference between DPT3 coverage reported by countries and estimated from data contained in Demographic and Health Surveys. It suggests that the official figures derived mainly from routine reporting systems over-report coverage in a large majority of countries compared to DHS as a gold standard (32).

Population surveys can sometimes be used to counteract the problems inherent in service registries. Even though they also have their limitations (costs, sampling errors, recall biases, interviewer bias, training of surveyors, etc.), they offer a valuable way of improving the validity, reliability and comparability of coverage data. Our strategy for measuring effective coverage in health systems is, therefore, to design a special module which will be incorporated into the World Health Survey, discussed in section VII.

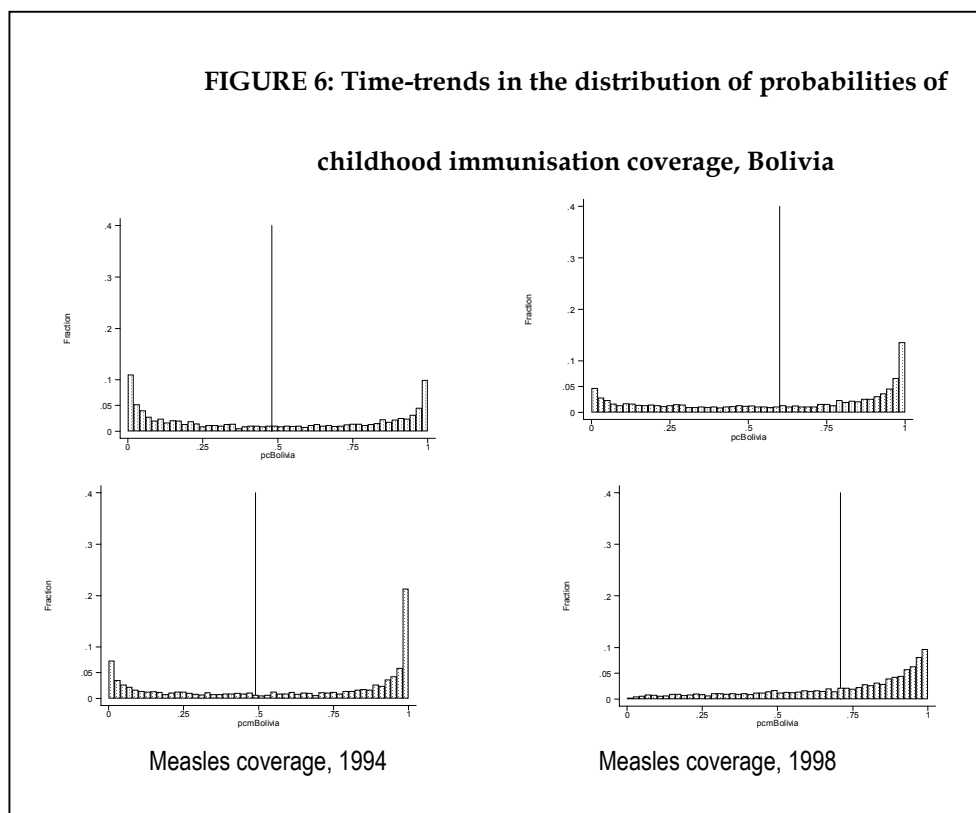
3. *Inequality in Effective Coverage*

The conceptual framework on coverage presented above has defined effective coverage at the individual level. In principle, it is possible to measure the distribution of effective coverage across individuals, something that is likely to influence inequalities in health outcomes. The distribution of effective coverage can be summarized using an inequality measure.

For example, we have estimated the probability of a child being covered with DPT3 or measles immunisation from empirical data contained in the Demographic and Health Surveys (DHS) in 46 countries, using a multivariate logistic regression model with a random-effects estimator. (33) Indices of inequality in coverage were then estimated for each country using an

individual-mean differences approach (that is, focusing on differences in the probability of coverage across individuals relative to the average level for the population).

Significant differences were found in the inequality of immunisation coverage across countries. Within countries, there were also differences in inequality between the two types of immunisation. Inequality also varied over time in countries where more than one survey was available. In Bolivia, for instance (see Figure 6), the spread of probabilities for immunisation coverage attenuated while at the same time the estimated average coverage level increased for both DPT3 (from 48% to 60%) and measles (from 63% to 72%). As a result, the inequalities index fell (became more equal), from 0.50 to 0.44 for DPT3, and from 0.47 to 0.34 for measles coverage.



4. Other Indicators of Service Provision

Other types of key indicators of service provision can also be defined. For example, in order to capture the contribution of the providers' actions to the outputs and outcomes of the health service provision process, it is important to understand provider performance. Provider performance measures the direct consequences of providers' actions on the population, a probable determinant of the effectiveness of the interventions that are delivered. It is not yet clear how much provider performance separately contributes to the attainment of intrinsic goals of health system, which is why we suggest including it here.

Indicators of the extent to which the structural organization and processes of different ways of providing interventions are likely to lead to effective coverage, thereby ensuring that the provision function contributes to health and responsiveness as desired, are more controversial. Three sets of variables could be proposed: (i) autonomy and integration, (ii) incentive structures and (iii) management of resources.

It is postulated that the level of **autonomy** of lower levels of the system and the degree of integration in the provision of services could increase the efficiency of service provision. However, the direction of the correlation is not clear. It has been suggested that the autonomy of service providers may improve technical and allocative efficiency, but when autonomy is not associated with appropriate incentives these potential benefits may not be realized. Further, autonomy may make it more difficult to obtain the benefits of economies of scale and scope.

The **integration** of health services is the process of bringing together common functions within and between organizations to solve common problems, by developing a commitment to a shared vision and goals, and by using common technologies and resources to achieve these goals. The evidence about the potential benefits or risks of integration is still not very convincing.

Incentive structures have an impact on the way providers of health interventions and actions behave, but there has been no clear agreement on what types of incentives maximize performance. We believe that it is worth defining, validating and measuring indicators which allow a variety of hypotheses to be tested.

An important component of **management** is the deployment of a required combination of different inputs. We propose to assess and monitor the management and deployment of inputs through: (i) the recurrent costs of service provision, (ii) the physical availability of inputs, (iii) the skill-mix of health care personnel, and (iv) utilization of medical equipment and structures.

Choosing indicators for autonomy, integration, incentive structures, management and provider performance requires consideration of whether they can be used to describe both personal and non-personal health interventions, whether each indicator provides additional information to that provided by the others, and whether it can be used to compare progress over time and across countries.

Proposal – Service Provision

WHO proposes that:

- 1. Effective coverage should be measured as a key instrumental goal of health systems. Effective coverage with a set of critical interventions defined for each country taking into account their epidemiological profile and health system development should be monitored. Tools for monitoring effective coverage should be developed, pilot tested and included in the World Health Survey**
- 2. Inequality in effective coverage should be measured and reported.**
- 3. A set of additional indicators related to the structure and organization of service delivery are also proposed for further exploration (Table 4). This list of proposed indicators is too long to be feasible. Suggestions on the most important and promising indicators to be measured would be helpful.**

Table 4. Proposed Indicators of the Service Provision Function

Concept	Question	Indicator	Potential Source(s)
Recurrent Expenditures	How efficiently are the inputs used in the production function, estimated through the magnitude and composition of recurrent expenditures?	The share of recurrent expenditures as a % of total health care spending	Facility surveys, National Health Accounts
		<i>The share of the wage bill as a % of total recurrent expenditures</i>	Facility surveys, National Health Accounts
		The share of expenditures on drugs and other medical supplies as a % of total recurrent expenditures	Facility surveys, National Health Accounts
		<i>Operating costs of structures and equipment as a % of total recurrent expenditures</i>	Facility surveys, National Health Accounts
Availability and utilization of drugs and other medical supplies	Are the necessary supplies available to deliver services and sustain the patient flow?	<i>The proportion of health care facilities that have essential drugs in stock</i>	Facility surveys
		<i>The proportion of generic drugs in the essential drug stock of facilities</i>	Facility surveys
Skill-mix	How efficiently different types of health care personnel are used?	<i>The ratio of different categories of health care personnel per 1000 first referral hospital beds</i>	Provider surveys, Labour force surveys, The Ministry of Health, The Ministry of Labour, Professional registries
Utilization of structures, medical equipment, and information technologies	How different physical resources are utilized?	Bed occupancy rate at primary, secondary and tertiary level facilities	Facility surveys
		<i>The proportion of medical equipment underutilized</i>	Facility surveys

Autonomy	Where the locus of decision making power lies and how it affects different management functions?	<i>The proportion of institutional providers who have full autonomy in human resource management and labour market issues</i>	Provider/Facility surveys
		<i>The proportion of institutional providers with an autonomous budgeting process</i>	Provider/Facility surveys
		<i>The proportion of institutional providers with the authority to independently contract out services</i>	Provider/Facility surveys
		<i>The proportion of institutional providers who can autonomously decide on the type and volume of services to provide</i>	Provider/Facility surveys
Integration	How well those tasks and functions are brought together that require the similar capacities, address the similar issue, and can benefit from economies of scale and scope	<i>The proportion of primary care facilities in which health services are fully integrated</i>	Facility surveys
Provider Incentives	How incentives facing health care providers determine their motivation?	<i>The proportion of health care providers by different modes of payment</i>	Providers surveys
		<i>The proportion of individual providers who hold both public and private sector jobs</i>	Provider surveys
		<i>Motivation of health care providers</i>	Provider surveys
Provider performance	How can providers' professional actions explain the outputs and outcomes of service delivery?		
		<i>Medical mistakes</i>	Provider surveys, population surveys
		<i>Hospital readmission</i>	Facility surveys, population surveys
		<i>Return to operating theatre</i>	Facility surveys, population surveys
		<i>Prescription patters</i>	Providers surveys, population surveys
Effective coverage	How well health systems meet health care needs of the population?	<i>The size of the realized health gain as a fraction of the total potential health gain possible if providers performed at the optimal level for a given health system</i>	Population surveys (WHO's World Health Survey)

D. Resource Generation: Human Resources and Capital

Health expenditures were used in the WHR2000 as the primary source of information on health system inputs for performance assessment. A number of the regional consultations suggested that it would also be useful to have information on the components of health expenditure, the inputs to producing the outcomes that people value. Human resources are particularly important because health systems are labour-intensive and expenditure on personnel is usually the biggest single item of recurrent health expenditure. Health systems require qualified and experienced staff to function well, and an appropriate mix between different types of human resources.

The quantity of health personnel depends on the decisions that the stewards of the health system make about investments in education and incentives designed to keep personnel in the system and to encourage or discourage migration. Decisions about accreditation and licensing influence the quality of health care providers, and salary and incentive policy influences motivation and provider quality. It is not, therefore, simply a matter of counting numbers to fully understand the potential contribution of human resources to goal attainment.

Similar comments are applicable to physical capital where deficits in the stock of assets (buildings and equipment, for example) can be a real constraint to the delivery of effective interventions in the short run. Conversely, it is not uncommon to find health systems where there has been significant investment in physical infrastructure but where recurrent budgets do not allow funding for the staff or the maintenance required to work the capital, resulting in efficiency losses. In the long run, investments are variable costs, and policy-makers need to carefully plan future investments in order to avoid shortages or excess capacities.

Goal attainment is influenced by the number and type of health facilities and the equipment available. Again, various types of incentives and legislation influence how capital is purchased, used and maintained, so it is not just a matter of counting numbers.

1. *Human resources*

Human resources can be defined as the stock of all individuals engaged in health actions - the promotion, protection, or improvement of health. This includes private and public sectors and different domains of health systems such as personal curative and preventive care, non-personal public health interventions, health promotion and disease prevention. For the assessment and monitoring of resource generation function we propose the following definition and classification of human resources (Table 5): (34)

Table 5. Definition and Classification of Human Resources

Type	Description	Challenges
Health professionals	Health professionals generated by the health care system in either full or part. Includes doctors, nurses, midwives, psychologists, pharmacists, dentists, and others.	The health system expects to employ 90% of the health care providers, which it generates. If this expectation is not met, shortages will follow. Competition for health care providers may be external – migration from developing to developed countries for example, or internal – from the public to the private sector.
Other professionals	Those workers of health care system who are not health professionals.	The health system must compete in the wider labour market to employ non-health professionals. In the UK, for example, managers were recruited to the health services from industry but to do so required a substantial raise in the salary levels of managers, in order to compete with salaries in other sectors

Three types of costs are associated with human resources for health systems: (a) investment costs spent on their production (capital expenditures on educational facilities, expenditures on training and education), (b) maintenance costs (continuing education), and (c) salaries and other benefits paid or offered to human resources.

The indicators in Table 5 provide useful information for planning and managing an effective human resources policy for health. (35) Some involve keeping track of the numbers of different types of personnel retained in the system, while others relate to the level of investment in new human resources. The indicator of motivation is the proportion of people who have been trained in the system, who are not at retirement age, and who remain working in the system.

2. Physical capital.

Physical capital is the stock of non-human productive assets including hospitals, vehicles, and equipment. Information on physical investment is essential for policy-makers as they make decisions about the allocation of resources now as opposed to investments that will support the provision of health services in the future. While it may be interesting to have a large number of indicators, this places considerable demands on health information systems so we propose a parsimonious set designed to provide the most important insights into the operation of the resource generation function.

They are found in Table 6 with full justification provided in the background documents.(34) The first 2 relate to the relative importance of investment in physical infrastructure compared to total expenditure and the funds available to maintain the infrastructure . The last simply provides information on the stock available to the system – a measure of input availability as described in section III.

We cannot be sure what proportion of total health expenditure is appropriate to devote to investment in physical infrastructure, nor what should be the ratio of maintenance to investment. Clearly, if too few operational funds are provided to allow health facilities to

operate effectively, there will be inefficiency but too high an allocation of operational funds will prevent facilities being available in the future. Accordingly, we propose to measure these indicators to help identify if there is any ideal level or proportion for investment in physical infrastructure.

3. Knowledge.

The stock of scientific and medical knowledge is a public good. Some can be obtained internationally, while knowledge specific to the application of tools and interventions must be obtained locally. We propose that expenditure on research and development (R&D) should be monitored yearly, broken down by public and private sectors. Again, we are unable to identify if there is an appropriate level of R&D expenditure, or whether more is better than less. But measuring it regularly and linking variations in this type of expenditure to variations in outcomes will contribute to exploring this question.

Proposal: Resource Generation.

WHO proposes to develop measurement strategies for the indicators proposed in Table 6, and begin to report on those where it is feasible and affordable to obtain reliable, valid and comparable measures.

Table 6. Proposed Indicators for the Resource Generation Function

Concept	Question	Indicator	Potential Source(s)
Generation of human resources	What types of human resources are available, and how much does their production cost to health systems?	<i>Total annual investments in human resources as a % of total health expenditure</i>	National Health Accounts
		<i>The ratio of the number of new graduates from health educational institutions over the total stock of health personnel.</i>	Medical schools, Ministry of Health, Ministry of Education
		<i>The total stock, composition, and distribution of human resources</i>	Provider surveys, labour force surveys, professional registries, censuses
	What is the pattern of human resource migration in country?	<i>The number of foreign health workers</i>	Provider surveys, labour force surveys, professional registries, special bodies of government dealing with immigration issues
Generation of physical resources	What is the stock of the physical assets and how much does it cost to the system?	<i>Annual new investments in health care facilities</i>	National Health Accounts
		<i>The total cost of maintenance</i>	Facility surveys, National Health Accounts
		<i>The total number and state of health facilities in the system</i>	Facility survey, registries of the Ministry of Health
Knowledge generation	What proportion of total health expenditure is spent on the generation of scientific knowledge and know-how?	<i>The total annual investment in research and development</i>	National Health Accounts, Ministry of Health, Ministry of Education

E. Stewardship:

Stewardship was defined in the WHR2000 as the careful and responsible management of the well-being of the population, the very essence of good government. Stewardship of the health system is the responsibility of government – usually through the Ministry of Health. This does not, of course, mean that the government needs to fund and provide all health interventions. It needs, however, to set the direction for both public and private sectors and ensure that the health system contributes to the socially desired intrinsic goals. How well or poorly a government executes its stewardship role can influence all aspects of health system performance.

Building on the work of other organizations as well as the recommendations of meeting of experts on the stewardship function held in September 2001, who has identified a succinct number of essential ingredients or ‘core domains’ that appear to constitute ‘good’ stewardship. Domains should be relatively well-defined, distinct areas of responsibility that are thought to help explain effective stewardship, and preferably, on which action can be taken. For each of them greater capacity should be considered desirable. Collectively, the better these are carried out, the more effective health system stewardship will be and the higher will be attainment of intrinsic goals. These domains, attributes and relationships are based on prevailing notions of effective stewardship. All should be considered ‘testable hypotheses’.

Six domains or sub-functions of stewardship have been tentatively identified (36)

- Generation of intelligence
- Formulating strategic policy direction
- Influencing behaviour through effective regulation
- Influencing behaviour through coalition building & communication
- Influencing behaviour by creating an enabling environment
- Ensuring accountability

The scope and core attributes of each domain are outlined briefly here and described in more detail in the background documents. Current thinking on approaches assessing stewardship are presented for comments from the peer review group. As yet we have not reached consensus on the indicators that can be used to measure the stewardship function.

- *Generation of intelligence.* This domain responds to the concern ‘to what extent do health system actors have useful intelligence at their disposal?’ do key actors have reliable, up-to-date information on current and future trends in health and different aspects of health system performance; important contextual factors and actors; possible policy options, based on national and international experience.
- *Formulation of a strategic policy framework.* This domain responds to the concern ‘to what extent is there a clear sense of vision and strategic direction for the health system?’ is there clear articulation of health system objectives; clear definition of roles and responsibilities of public and private actors in all four functions; identification of policy instruments and institutional arrangements required to achieve improvements, and strategies for making the changes; guidance for prioritising health expenditures based on

realistic resource and needs assessment, and for monitoring effects of changes on performance?

- *Influencing behaviour through effective regulation.* This domain addresses the concern 'to what extent is there a regulatory framework that facilitates implementation of health policy, i.e. steers different actors in the desired direction?' 'regulatory framework' refers to the spectrum of rules, procedures, laws, codes of conduct, standards etc that exist. This will involve looking at the scope of existing regulation; conflicts / contradictions between stated policy and the regulatory framework (whether powers and responsibilities are matched); the extent to which they are enforced.
- *Influencing behaviour through coalition building and communication.* This domain addresses the concern 'to what extent does capacity exist to create alliances of individuals, groups or organisations for joint action around strategic health and health system priorities?'
- *Creation of an enabling working environment.* This domain addresses two questions. To what extent do organisational structures and management systems fit with policy objectives so that they help rather than hinder policy implementation; to what extent have conditions been created by government that allow stewards themselves to be effective?
- *Accountability.* This is considered a separate domain at present on the grounds that it is a stewardship responsibility to ensure that all health system actors are held accountable for their actions. This will also contribute to consumer protection.

Further work is needed to more rigorously map out the attributes of these different domains, before developing specific indicators. However, some thought is already being given to how one might obtain information on this function. Unlike financing, there are few existing routine data sources. Based on WHO's recent experience with measuring responsiveness, one approach that is being considered is to develop a survey instrument that would include questions on all domains of stewardship, accompanied by vignettes, that could be administered to all main groups of health system actors, and perhaps also from households.

Proposal - Stewardship.

WHO proposes to develop an instrument to describe the stewardship function, based on six domains - generation of intelligence; formulating strategic policy direction; influencing behaviour through effective regulation; influencing behaviour through coalition building and communication; influencing behaviour by creating an enabling environment; and ensuring accountability. The validity, reliability and comparability of the instrument should be established through appropriate psychometric evaluations.

V. HEALTH SYSTEM CONTRIBUTION TO SOCIAL GOALS

This section reviews the five components of intrinsic goal attainment outlined in Table 1 as well as a composite index of overall goal attainment. Each subsection will be organized according to the following components:

- a description of the WHR2000 measurements;
- commentaries on those measurements;
- new analysis undertaken to explore some of the suggestions and criticisms;
- proposals for the next round of performance assessment, for consideration by the Peer Review Group.

A. Average Level of Population Health

Health has always been measured and reported routinely by WHO. This process has benefited from interaction with Member States and with technical experts, the most recent being the Conference on Summary Measures of Population Health held in 1999, and a Meeting of the Committee of Experts on Classification and Measurement of Health in 2000.(22) Population health monitoring is a core constitutional function of the World Health Organization. It is not strictly an activity of health systems performance assessment. Nevertheless, population health measurement is an important input to health systems performance assessment. In this section, to facilitate the deliberations of the PRG, we provide an update on the current status of work by the WHO on population health measurement.

1. WHR2000

A number of indicators of population health have been reported by WHO for each Member State for many years. These include child and adult mortality risks and life expectancy at birth. In addition, WHO has been involved in the development of summary measures of population health for over a decade. Summary measures combine information on mortality and non-fatal health outcomes to represent population health as a single number (37) and WHO's involvement began with participation in the Global Burden of Disease studies in the early 1990s.(38) More recently, WHO has built on measures such as active life expectancy (39) and disability-free life expectancy(40-45) by proposing the concept of disability-adjusted life expectancy.(46) Measurement methods and data sources have been continually refined since then, and the WHR2000 reported disability-adjusted life expectancy (DALE) at birth and at age 60, for males and females separately. Uncertainty intervals around the most likely estimate for each country were also reported.

As part of this process, new life tables and life expectancies were estimated for all 191 Member States. For countries without adequate vital registration data or surveys, estimates were based on regional logit models.

To form the summary measure of population health for each country (DALE) it was necessary to supplement the life table data with information on the age and sex-specific prevalences of non-fatal health outcomes, and the appropriate health state valuations. Health state valuations were estimated for each major non-fatal health outcome for five standard age groups, by sex, in eight regions. Where the data were available, detailed information on the epidemiology of the major conditions in countries was used to construct prevalence of non-fatal health outcomes, along with the earlier Global Burden of Disease estimates. We tried to adjust these estimates with information on self-reported health from surveys undertaken in a number of countries, but found that most of these data were not comparable across settings. Full details were described in Mathers et al. (47;48) and Sadana et al. (49)

2. Commentaries on population health measurement presented in the WHR2000

Most of the criticisms emerging in the published literature and at the consultations is self-explanatory and is only briefly described here. More detailed information is provided in the background documents. The first set of debates focused on the life tables where it was argued that more vital registration data were available in countries than was used by WHO and that the logit model did not perform well at older age groups.

The second set concerned DALE. Of major concerns was the fact that existing surveys available to calculate the prevalence of non-fatal health outcomes produced measures that were not comparable across countries, across populations within countries, or over time. In addition, it was suggested that the indicator added little additional information to that contained in life expectancy because it was highly correlated across countries with life expectancy. Another concern was the fact that DALE would change only slowly over time so it was not of short term interest to policy-makers. Other questions were of a more technical nature and included:

- DALE was not an appropriate name because it really measures the time that a person can expect to live in equivalent good health rather than the extent of disability;
- Estimates of the prevalence of non-fatal health outcomes and their associated valuations had not been derived from an explicit multidimensional description of health but were based on defined diseases or conditions;
- Health state valuations used in the calculation of DALE were not based on representative household surveys.

3. New Analysis

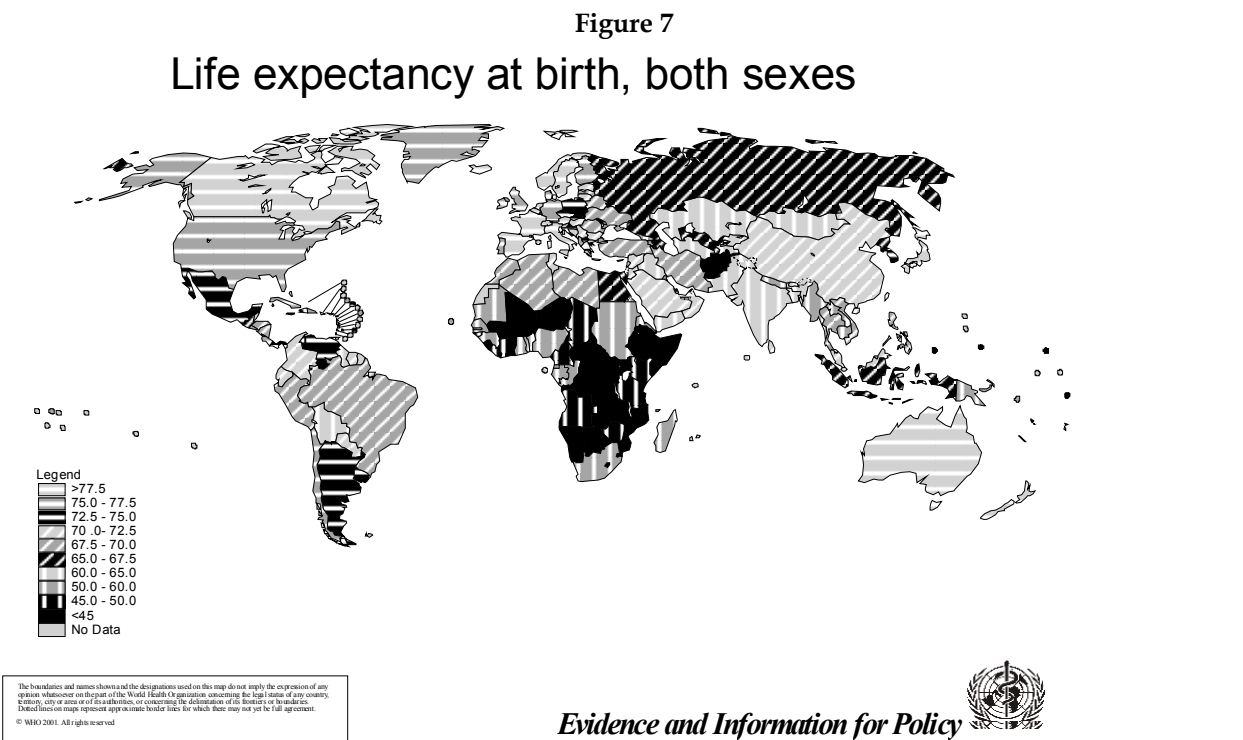
a) Life Tables

A new method of estimating life tables correcting for problems at older ages has been developed using the Brass-Logit system. This was used for the WHR2001 and has been validated by out-of-sample predictions.(50;51) In addition, intense interaction with regional and country offices of WHO and with Member States has provided information on vital registration in 128 countries compared to the 82 for which data were previously available. (Table 7)

Table 7: Countries where Complete or Partial Vital Registration Data are available

	Number of countries with recent Vital Registration data	
	WHR2000	WHR2001
Africa	3	7
The Americas	21	33
Eastern Mediterranean	3	10
Europe	43	51
South-East Asia	4	6
Western Pacific	8	21
World	82	128

New estimates of the distribution of life expectancy at birth by country and shown in Figure 7. New-born males in Sierra Leone could expect to live only 37.0 years on average whereas new born females in Japan could expect to live to almost 85 years in 2000.



b) DALE name

We have found that the concept of DALE is simple to understand, being based on the widely used concept of life expectancy. To reflect the fact that it is, indeed, a measure of how long people can expect to live in equivalent good health, we changed its name to HALE or healthy life expectancy early in 2001. It was reported as such in the WHR2001.

c) Cross-population comparability

One of the most complex problems over the last year has been to develop methods to ensure cross-population comparability of survey data. The way people classify their own health on different domains varies consistently with factors such as education, sex, age, and country. This is because of response category cut-point shifts, and this implies that self-report categorical data are not comparable across individuals. They cannot be used to infer the underlying levels of health without adjustment.

The goal of adjustment is to ensure that the same level of health results in the same measurement, irrespective of education, sex, age etc. WHO has developed novel strategies to ensure cross-population comparability. (52-54) They included development of survey instruments which included vignettes, as well as undertaking measured tests on some survey respondents. They included development of statistical techniques using the hierarchical ordered probit model.(55) The effect of using vignettes and the hierarchical ordered probit model is illustrated for the domain of mobility in Figure 8. In this survey in Colombia respondents were asked if over the last 30 days they had any difficulties moving around.

Response options were no difficulty, mild difficulty, moderate difficulty, severe difficulty or extreme/unable. The circles show the mean response by age. There is only a slight decrease in average mobility with age. Once the way in which respondents use the categorical scale is measured, the drop-off in mobility with age is more pronounced. A full description of cross-population comparability is found in section VII.

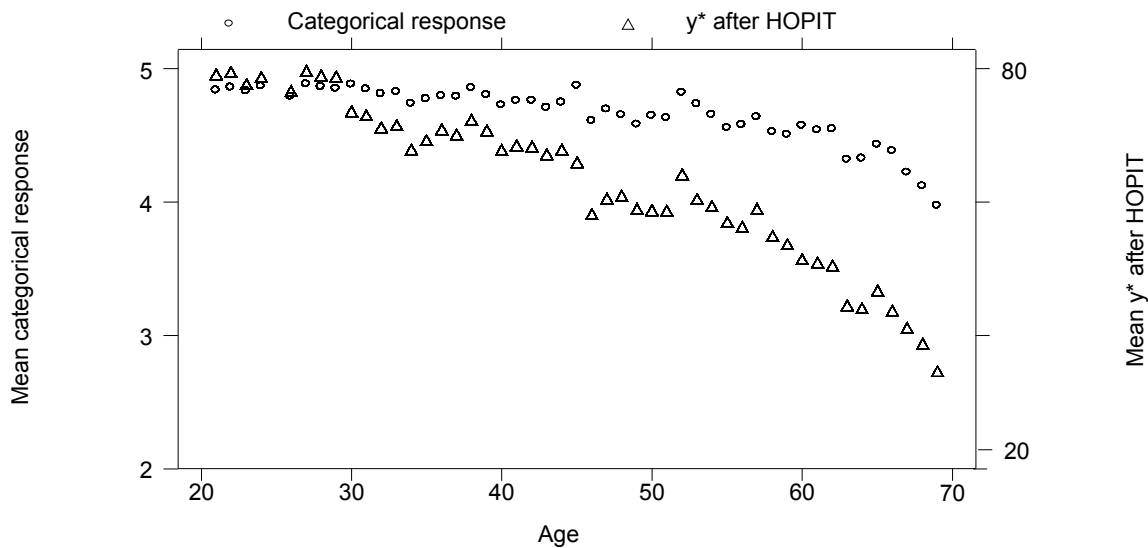


Figure 8. Unadjusted and Adjusted Mobility Scores by Age

d) Prevalence of non-fatal health outcomes

To address the multidimensional nature of health, a WHO Multi-Country Survey Study was undertaken in 61 countries from 2000 – 2001. All involved representative household or postal surveys and asked respondents to rate their current health according to 6 key domains. The domains were based on the International Classification of Functioning, Disability and Health which was ratified at the World Health Assembly in 2001.(56) More detailed surveys in 10 countries asked about another 15 domains. The survey instrument included the vignettes developed specifically for the purpose of establishing cross-population comparability. The study is described in more detail in section VII.

At the same time, updated estimates of the Global Burden of Disease (GBD) – disability adjusted life years lost by age, sex and cause – were being made for 14 sub-regions of the world. These estimates also provide regional estimates of prevalence of different non-fatal health outcomes. A three stage procedure was used to estimate the final prevalence of non-fatal health outcomes – updated Burden of Disease estimates for 2000, health survey data, and statistical methods which allow us to construct posterior estimates of HALE that maximise cross-country comparability.

For the 2001 estimates, we calculated posterior prevalences of health states from the GBD-based estimates and the health survey data, assuming that the uncertainty in both was normally distributed. We then used the relationship between the GBD-based estimates and the posterior estimates for survey countries (i.e. the overall pattern of change introduced by the survey data) to revise the GBD-based estimates for non-survey countries. This ensured maximum comparability across all WHO Member States. Figure 9 shows the age-standardised posterior prevalences for females for the 191 WHO Member States plotted against life expectancy at birth.

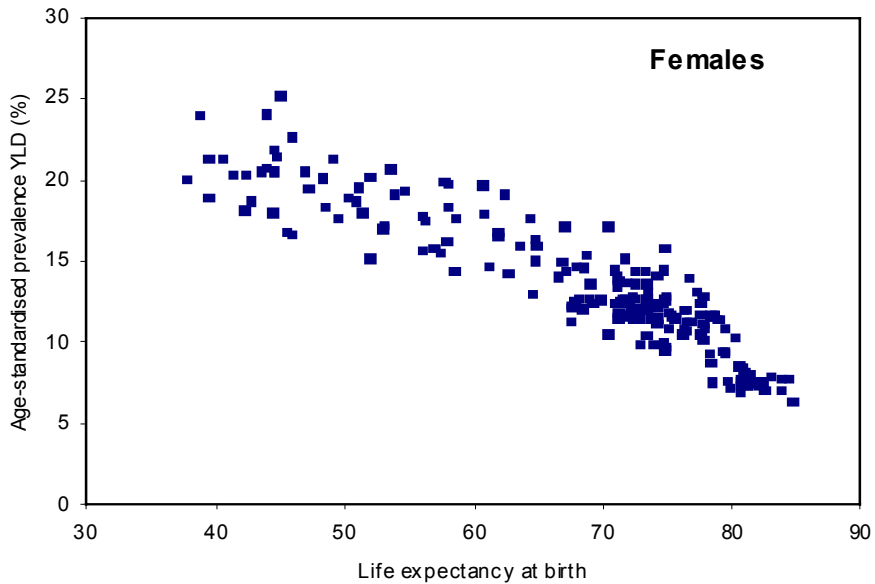


Figure 9. Estimated age-standardised posterior severity-adjusted prevalence rate versus life expectancy at birth, by sex, WHO Member States, 2000.

e) Health state valuations

Survey results on individual levels on key domains of health were used to estimate the non-fatal health state prevalences for HALE by applying a valuation function derived from the full-length household surveys in 10 countries. The valuation function constitutes an empirical means of mapping between multiple domains of health and a scalar index of the level of health measured on a cardinal scale anchored by perfect health and death. The valuation function was estimated based on individual descriptions of domain levels associated with 34 hypothetical health states and valuations of these hypothetical health states using methods described in further detail in the background document.⁶⁽⁵⁷⁾ A global valuation function was applied to the individual data on domain levels; examination of valuation functions estimated at the country level revealed minimal differences across the 10 survey countries.

Figure 10 shows the mean visual analog scale (VAS) results by country for each of the health states provided as stimuli to the respondents. For ease of exposition, every second health state is labelled. While there is some variation for any given state across countries, there is substantial consistency across countries overall. It is also important to note that in this study, individuals were asked to provide their own descriptions of how they imagined each state in terms of the core domains of health, so that in fact, the label “QUA” (quadriplegia) represents not a single defined set of domain levels across all respondents but a range of different states depending on how each respondent described quadriplegia.

⁶ Each individual was asked to value only 10 states plus their own health, but by rotating sets of health states across respondents we have data on 34 health states as well as own health.

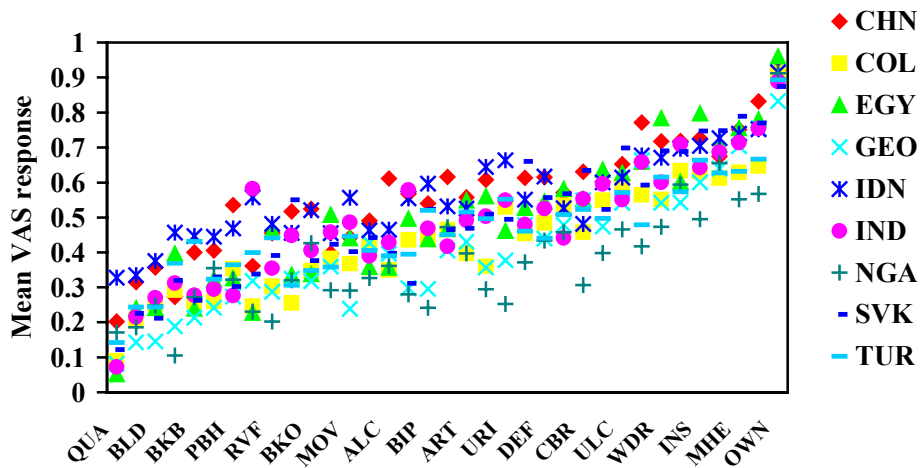


Figure 10. Average VAS scores by country and state⁷

We may, therefore, decompose the variation in visual analog scale responses across countries in terms of the variation due to different domain descriptions and the variation due to different valuation functions through which individuals map between a vector of domain levels and a scalar valuation of health. In Figure 11, the predicted ranks of each health state are plotted by country using two different sets of estimates. In the first panel, the visual analog scores by country are estimated using country-specific valuation functions applied to the average domain level descriptions for each state by country. In the second panel, the scores have been estimated using the country-specific valuation functions applied to the *global* average descriptions. The difference in the amount of variance between the first panel and the second illustrates that most of the variance in health state valuations across countries is actually due to variance in the individual descriptions of each health state rather than variance in the valuation function. This result supports the use of a global average valuation function.

7

Country Code	Condition Code	
CHN - China	QUA - quadraplegia	ART - arthritis
COL - Colombia	BLD - total blindness	URI - urinary tract infection
EGY - Egypt	BKB - below the knee amputation in both legs	DEF - deafness
GEO - Georgia	PBH - paralysis in both hands	CBR - chronic bronchitis
IDN - Indonesia	RVF - recto-vaginal fistula	ULC - pain in stomach, as in ulcer
IND - India	BKO - below the knee amputation in one leg	WDR - watery diarrhoea
NGA - Nigeria	MOV - movement disorder	INS - insomnia
SVK - Slovakia	ALC - alcohol dependence	MHE - mild hearing problems
TUR - Turkey	BIP - bipolar depression	OWN - overall health

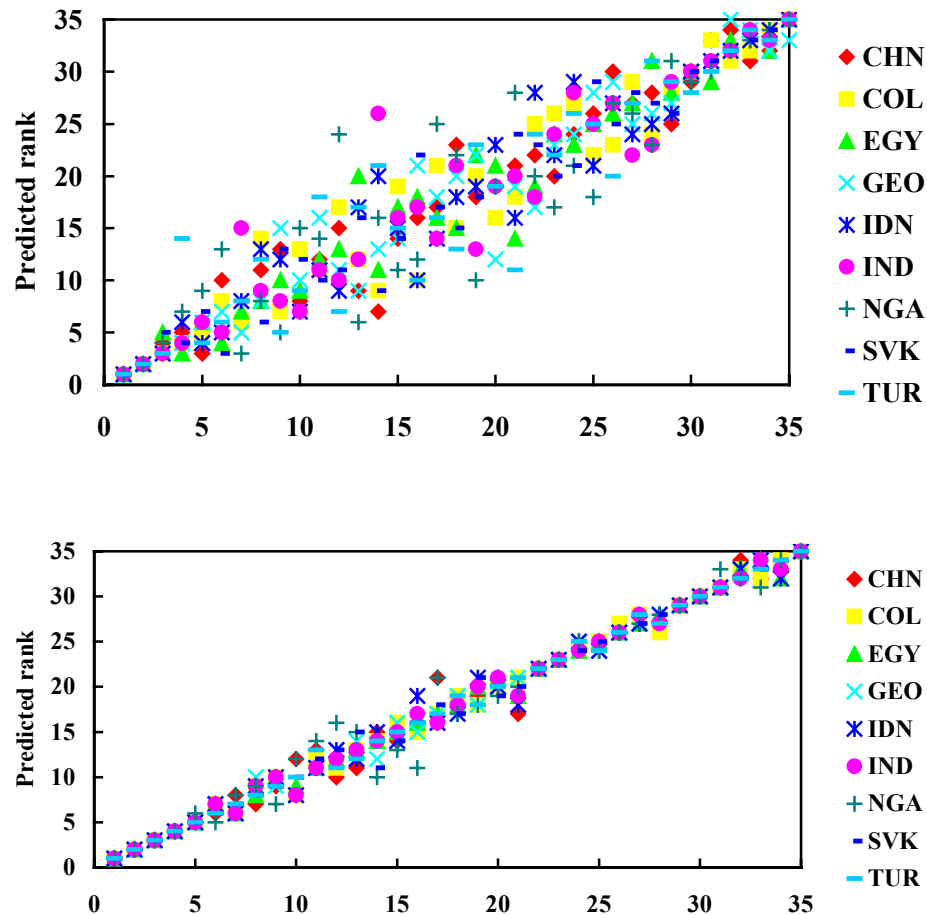


Figure 11. Predicted valuation ranks using country-specific valuation functions applied to country-specific domain descriptions (top panel) and country-specific valuation functions applied to global domain descriptions (bottom panel).

f) Correlation between HALE and Life Expectancy

The new estimates of HALE for 2000 suggest that in a number of countries, new-borns could expect to live more than 70 years in equivalent good health. On the other hand, children in some countries in Africa where the impact of the AIDS epidemic is the most severe could expect to live less than 35 healthy years.

HALE at birth is correlated with life expectancy at birth. However, it provides considerably more information. To illustrate, Figure 12 is taken from data reported in the WHR2001 which uses the new life tables, burden of disease and health surveys. The correlation across countries is high, with a simple correlation coefficient of 0.98. Despite this, the difference between life expectancy and HALE (measured on the vertical axis) can be very substantial at any given level of life expectancy (on the horizontal axis). For example, at a life expectancy of approximately 70 years, the HALE is only 5 years lower than life expectancy for males in some countries, but it is as much as 16 years lower for females in other countries. In addition, the gap between HALE and life expectancy increases up to a life expectancy of 68 for males and 72 for females. Above that level of life expectancy, the gap between HALE and life expectancy declines. This is dramatic evidence of the compression of morbidity hypothesis (58;59). One

explanation for the apparent compression of morbidity may be that health systems in high income countries may do more to reduce ill-health or the severity of ill-health than reduce age-specific mortality rates. Shifting from examining life expectancy to HALE may substantially alter our perception of the effectiveness of health systems. Clearly HALE provides considerably more information to policy-makers than life expectancy despite the apparent high correlation between the two.

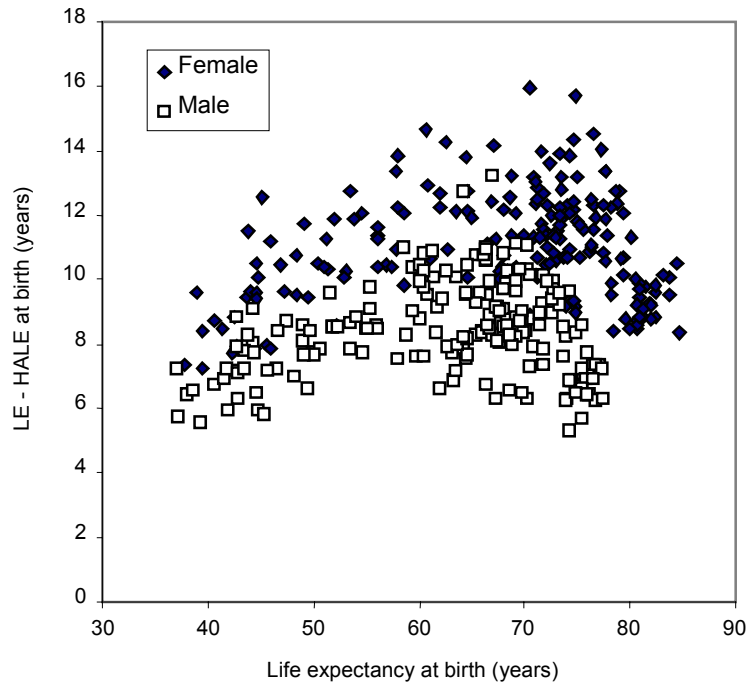


Figure 12. Life expectancy minus healthy life expectancy at birth plotted against life expectancy at birth, by sex, 191 WHO Member States, 2000

4. Proposals

The question of whether HALE will change rapidly over time is an empirical one. Certainly it has fallen dramatically in countries where HIV/AIDS is endemic and it fell for other reasons for males in some of the eastern block countries after the break up of the former Soviet Union.

We suggest that it could rise rapidly in some of the poorer countries if there are substantial injections of resources for health and if those resources are used efficiently. In the richer countries, it will rise less rapidly and additional process or instrumental indicators are useful to provide short term information to decision-makers. This is no different to life-expectancy. The fact that it changes slowly over time in rich countries does not prevent the primary quantity of interest being routinely calculated, reported and used by policy-makers.

WHO proposes that for the particular needs of health systems performance assessment, further work on the measurement of population health be focused on:

1. **Develop a standardised incidence-based HALE that extends the concept of a period measure to estimate the expected healthy years of life for a new born child who experiences at each age the current incidence (rather than prevalence) of health problems and the current patterns of exposure to risk factors. Such an incidence-based**

HALE would allow the measure of health outcome to reflect current population risks and health system activities directed at such risks, particularly those which are new or have long lag times to health outcomes, rather than population risks associated with past exposures. To ensure comparability, it will be important to specify standard assumptions for the relationship between risk exposures and outcomes, and to select a standard set of factors to be taken into account in the analysis. This set of factors may be expanded in future years to include additional risk factors or interventions as methods and data improve.

2. To develop an incidence-based HALE that would be sensitive to the currently observed prevalence of key risk factors, valid, reliable and comparable data on key risk factors would need to be collected. One promising strategy to obtain this information would be to include a risk factor module in the World Health Survey. Since the first global burden of disease exercise was published concern has been expressed that most information on the major causes of ill health relates to disease or conditions. Not enough is known about risk factors and the World Health Survey offers the appropriate platform for its collection and subsequent analysis.

B. Health Inequality

5. WHR2000

To be logically consistent with measuring health using healthy life expectancy, WHO outlined the intention to assess inequalities in HALE. (60) We began by measuring inequality in the probability of child survival for the WHR2000. This is a critical determinant of HALE, particularly in developing countries. A parametric model to estimate the distribution of the probability of surviving to age 2 was used. This distribution was summarized with a measure of inequality, similar to those used in the economic literature to summarize the distribution of income.

The beta-binomial model was used to distinguish between variation across mothers in the number of children who died due to chance, and the variation due to differences in the underlying risks of death. Data were taken either from birth histories available in demographic and health surveys or from child mortality variations across small geographical areas - 56 countries had detailed micro data. The distribution of the risk of child death was transformed into a distribution of expected survival time to age 5 years, and a composite inequality index of child survival was defined as:

$$II[3,.5] = \frac{\sum_{i=1}^n \sum_{j=1}^n |h_i - h_j|^3}{2n^2 \bar{h}^{0.5}}$$

where h is the expected survival time of a given child and \bar{h} is the mean survival time for all children in that country. This measure was transformed into an equality measure, simply by taking $(1-II)$. The parameters used for this summary measure were based on an informed preferences survey of over 1600 respondents.

For the countries where data were available, a statistical model was applied which estimated the relationship between equality in child survival and other available covariates. This model was then used to estimate the index for countries where data were not available. Uncertainty intervals around the most likely estimate were reported for all countries.

This is an index which relates each individual observation to all other individual observations while taking into account the mean as well. The particular functional form of the measure heavily weights the tails of the distribution.

6. Commentaries on the WHR2000

One philosophical and a number of technical questions have been raised in the regional consultations, a technical consultation and the academic literature. The normative issue is whether pure health inequalities should be measured as in the WHR2000 or whether the only interesting measure is average differences in health by social class or income groups.

The technical questions raised can be grouped into three general categories. First, some argued that there were too few countries with surveys or small area data, and too many where the equality index was estimated from covariates. Some commentators were concerned that child mortality is only one component of health, so inequalities in child mortality capture only one source of health inequality. Third, the issue was raised that the index is not bounded on a 0-1 scale, and could, at least theoretically, be negative.

7. New Analysis

a) Sources of inequality

As with the literature on income inequality, the WHO approach makes it possible to analyse the determinants of child survival inequality. This approach allows for the quantification of the existing inequalities in health and the exploration of all potential determinants, (61) rather than focusing on one or two factors which are believed *a priori* to be correlated with the outcome. To explore this question, the child survival equality index has been decomposed into the contribution of different factors. The Demographic and Health Surveys provide information on some socioeconomic variables such as household assets and mother's education, and several variables that together can be considered as a proxy for access to the health system – including antenatal visits, measles immunization, BCG immunization and births attended by a nurse or doctor.

To study the effects of these covariates a counterfactual analysis was performed in which the inequality index was recalculated assuming first that there were no inequalities in assets, then that there were no inequalities in education and third that there were no inequalities in access to the health system. Because it is a critical question if health system actions can reduce health inequalities, the impact of expanding health system access on child survival inequality has also been assessed.

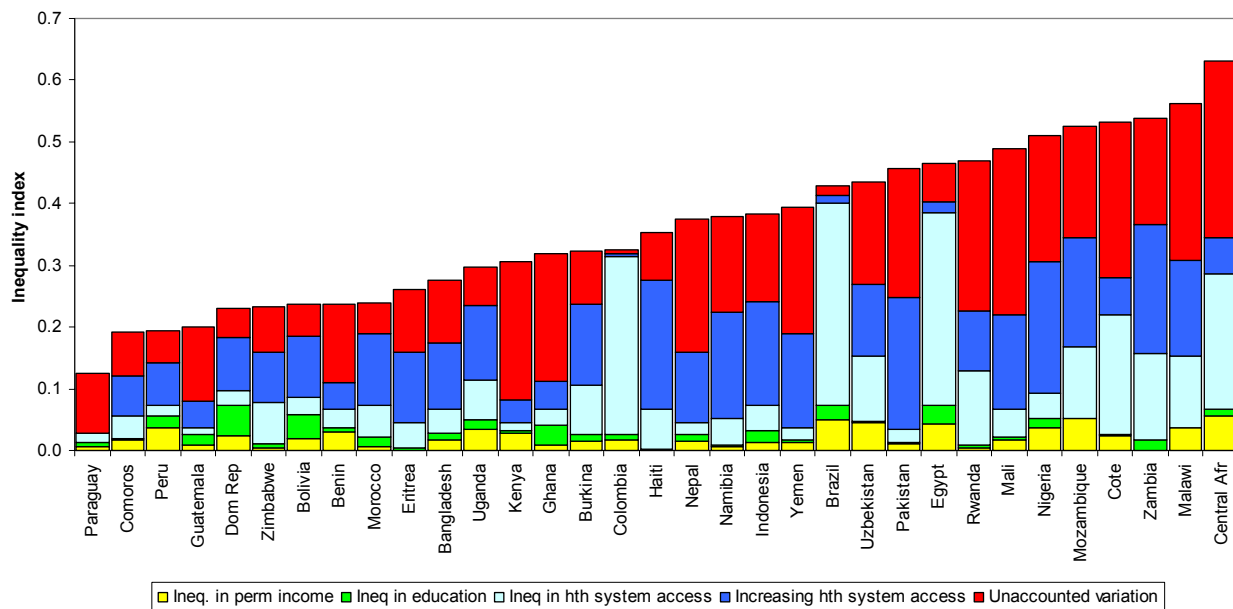


Figure 13. Decomposition of Child Survival Inequality

This analysis suggests that reducing inequalities in assets and education would reduce inequalities in child mortality to some extent in all countries (Figure 13). Reducing inequalities in access to health systems would also reduce inequalities in child survival in all countries. On average the contribution of inequalities in access to health systems to inequalities in child survival is larger than the contribution of reducing income or education inequalities.

It is even more encouraging that providing greater access to health services (to the extent that our proxy variables reflect access) could have a dramatic effect on reducing health inequalities and lowering the average level of child mortality. This is, of course, exploratory and simply suggests that if we had focused only on assets or education, we would have missed the chance to consider other possible determinants. This does not imply that policy makers are not interested in measuring inequality in health across socioeconomic groups. They are. But it also implies that measuring pure inequalities in health allows disaggregation by all possible determinants, including socio-economic ones.

b) HALE in the Poor

Because of the keen interest in improving the health and well-being of the poor, WHO is developing methods for estimating HALE in the poor. This could then be compared to average HALE in a population as a complementary indicator to inequality in HALE at the individual level. In countries with linked individual data on mortality and various individual covariates including household income, it is relatively straight forward to calculate and report HALE in the poor. Unfortunately, at present linked individual datasets are only available for a small number of countries.

WHO would like to develop more widely applicable strategies to measure HALE in the poor. This requires a valid, reliable and comparable estimator of permanent income to be developed. A new module is being added to the forthcoming World Health Survey aimed at measuring permanent income as a first step. Application of the hierarchical ordered probit model (55) to asset and permanent income indicator questions has demonstrated that valid estimates of permanent income can be obtained. For each country, it is possible to map the level of permanent income at which a household is likely to have an asset or a particular indicator, such as running water or electricity. This information will make it much more feasible to report on levels of health in the poor in a wide range of countries.

c) Technical Issues: Child Mortality

Since the WHR2000, an additional 17 demographic and health surveys and 13 surveys from the Pan Arab Project on Children have been made available to WHO. WHO expects to be able to analyze between 30 and 60 data sets from the Multiple Indicator Cluster Surveys (MICS) implemented by UNICEF. Relevant birth history data from up to 120 countries should be available. This will allow greater precision in the estimation of inequality in child mortality.

d) Technical Issues: Adult Mortality

A survival analysis model has been developed to be used in the estimation of the distribution of adult mortality risk (62). This model is similar to the one used for children and can be used on individual level data which are available for a number of Member States from health surveys and censuses that have been linked to vital registration data. The survival analysis model includes a shared frailty component that can capture unmeasured community effects on adult mortality. Figure 14 shows one application of this model to data from the USA using the distribution of mortality risk for males and females aged 45-54 years. As expected, the

average mortality risk is lower for females than for males and their also appears to be less variation in this risk for females.

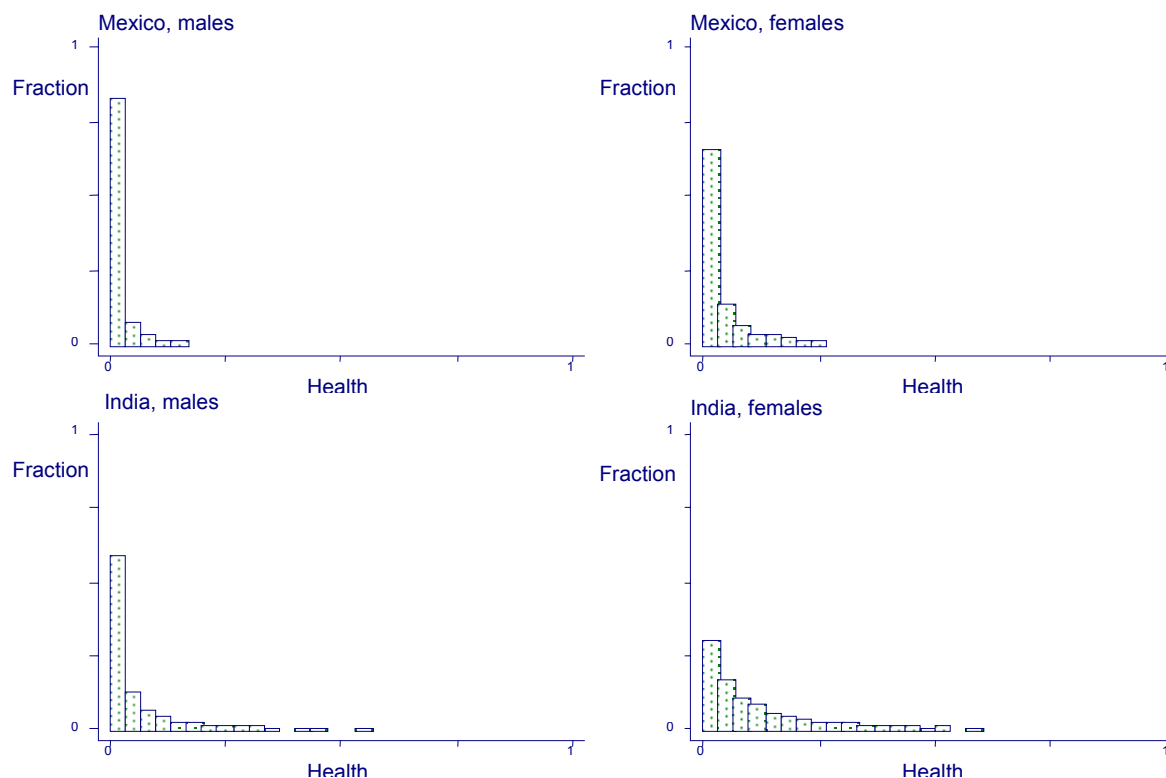


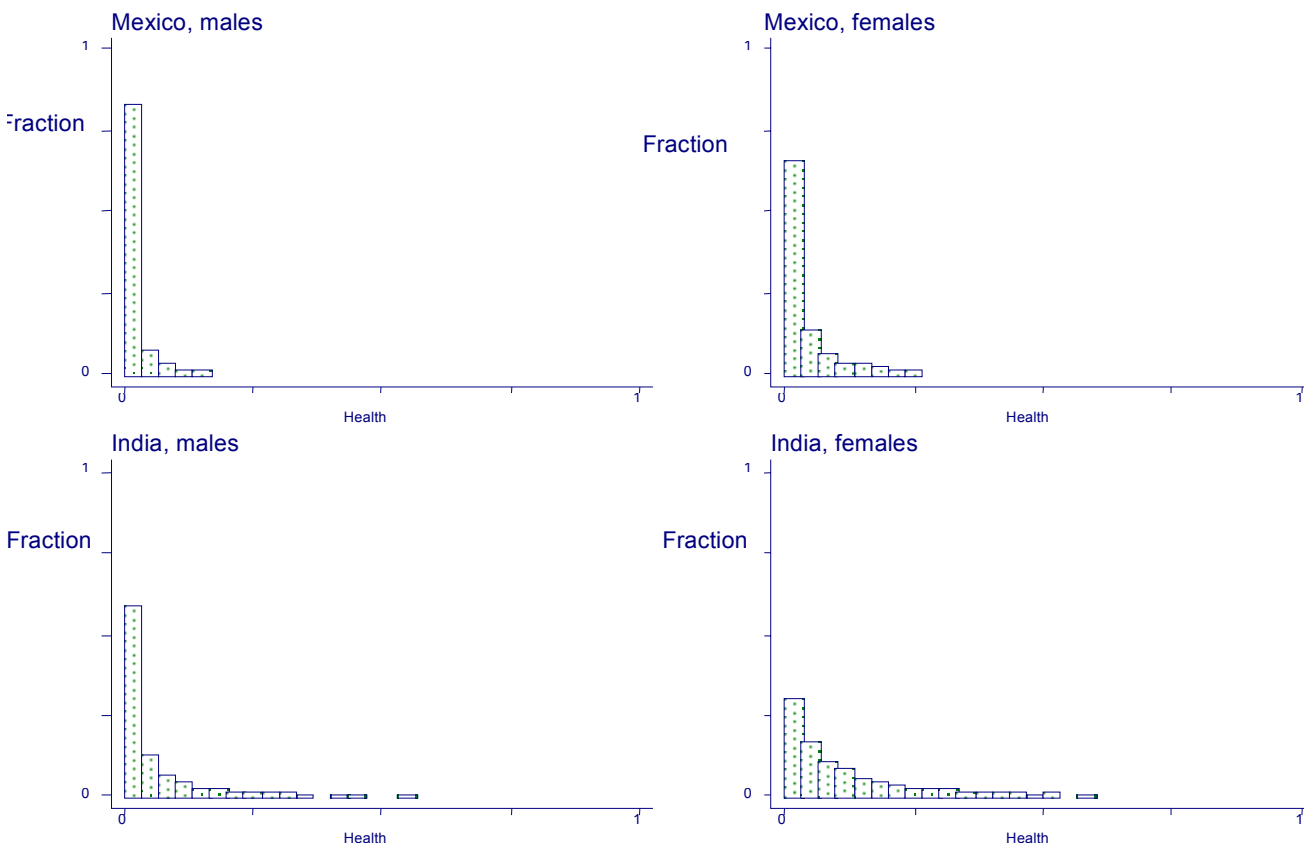
Figure 14. Predicted distributions of risk of death for males and females, aged 45-54 years, US.

For Member States where individual-level data are not available, statistical techniques have been developed to estimate the distribution of mortality risk from small area data. Inevitably, small area data will underestimate the true level of inequality. To estimate the extent of under-estimation, results from small-area data will be compared to results from individual-level data for about 10 Member States where both types of data are available. If a systematic relationship is found between the estimates from the two types of data, it will indicate that small area data can be used to estimate the distribution of mortality risk for adults where individual-level data are not available.

e) Technical Issues: Non-Fatal Health Outcomes

Inequality in child survival is clearly more relevant for low and middle income countries and the goal is to extend the approach to adults mortality and non-fatal health outcomes. One component of this work is the quantification of the distribution of health states across individuals using data from the 2000-01 WHO Multi-country Survey Study. Preliminary results of variability in non-fatal health outcomes across countries included in the survey study are found in Figure 15. Here, the distribution of health states in for adult males and females (aged 15-59 years) in India and Mexico are shown. Results should be interpreted with care because of the different modes of surveys that were included, but they show the power of using survey data to estimate the variability in the prevalence of different non-fatal health outcomes. Figure 15 indicates that there is more variation in non-fatal health outcomes in adult females than adult males, which is different from the observed trends in risks of mortality, where there is more

variation in males than females. These preliminary results highlight the importance of measuring inequality in HALE, a measure of health which includes risk of mortality and non-fatal health outcomes.



f) **Technical Issues: Measure of Inequality.**

The discussion around the measure that should be used to summarize the quantity of interest was not conclusive in the technical discussion. However, there was general agreement that a combination of measures may need to be calculated to encompass concerns about inequality. It was agreed that WHO does need to use a single measure in its final estimation of health system attainment and that such a measure need not reflect the average level of health, as that is reported separately by WHO. One suggestion was that a new survey could be fielded by WHO to measure preferences for a measure of inequality.

Proposals – Health Inequality

WHO proposes:

1. **Measuring pure health inequality and average levels of health across socioeconomic groups are complementary rather than competitive approaches.**
2. **Three indicators of health inequality be measured and reported routinely:**
 - **child survival inequality, based on the new data outlined above and a new random effects logit method described in the background documents;**
 - **HALE inequality; and**
 - **HALE of the poor.**

3. **Measurement methods using data from the World Health Survey, small areas and individual-linked datasets should be further developed, tested and applied to measure HALE inequality and HALE of the poor.**

C. Fairness of Financial Contributions

1. WHR2000

The fairness of financial contribution to a health system starts with the question: taking society's efforts to redistribute income as a given, what is a fair contribution to the health system? As a normative claim, the WHR2000 proposed that the sacrifice created by contributing to the health system should be equalized across households independent of their health status or utilization of health services. This equal sacrifice can be transformed to an equal share of all households' capacities to pay. This implies that the goal of the health system is not to redistribute income, but people want their health system to be financed in a fair manner.

Household payments to the health system included all financial contributions attributable to the household through taxes, social security contributions, private insurance, and direct out-of-pocket payments. Household capacity to pay was defined as household effective income net of subsistence expenditure, where effective income was the level of consumption that a household would seek and is able to consume, based on a life-cycle perspective. Subsistence minimally included expenditure on food, basic shelter and minimal clothing but not health. For the WHR2000, capacity to pay (CTP_i) was measured as total consumption expenditure minus food consumption and household financial contributions (HFC_i) were total contributions to health divided by capacity to pay.

The fairness of financial contribution index measured the distribution of household financial contributions. Since catastrophic spending is of particular concern, the summary index of the distribution of household financial contributions was designed to give heavy emphasis to the right tail of the distribution.

The formula used to calculate the Fairness of Financial Contribution Index was

$$FFC = 1 - 4 \left(\frac{\sum_{i=1}^n |HFC_i - \overline{HFC}|^3}{0.125n} \right)$$

where

$$\overline{HFC} = \frac{\sum HFC_i}{n}$$

This is an index of individual-mean differences rather than an index of inter-individual differences. It is built on the assumption that people care about their place compared to the average contribution in the society and judge inequality accordingly. The second family assumes that people mostly care about their place compared to other individuals, unrelated to the average in that society. The choice was made after conducting an internet survey of over 1,600 people in which questions were constructed to elicit individual preferences for equality. The results showed that with respect to financial contributions to the health system a large

majority of people cared more about the difference between their contribution and the average level of the whole population than the difference between their contribution and that of all other individuals in the population.

Household income and expenditure surveys which were recent and which had the appropriate data for construction of the index were found in 21 countries. The results of those countries were used in a regression analysis to identify critical covariates of the FFC index and this regression was used to estimate the FFC for countries where household surveys were not available.

2. Commentaries on the WHR2000

One of the main concerns with the FFC was that it was based on micro level analysis of income and expenditure surveys in only 21 countries with regressions used to estimate the values for the remaining countries as a function of covariates such as the fraction of total health expenditure from out-of-pocket payments. (63;64)

A summary of other major comments and suggestions follows.

- The FFC as defined could penalize countries with very progressive payment systems because in these countries the rich may pay a larger share of their capacity-to-pay than the poor, even allowing for the subtraction of subsistence expenditure from income in the definition of capacity to pay. In these cases, the distribution of HFC could be more unequal and the FFC index lower.
- Food consumption is an inadequate proxy for subsistence expenditure because much of the food consumption of the rich is not for subsistence needs. Subtracting all food consumption from total income underestimates the capacity-to-pay of richer households.
- The FFC does not take into account the utilization of the service. When the poor spend less it might be because they cannot afford care at all.
- The interval scaling properties of the FFC had not been established.
- Policy-makers may be interested in the income redistributive effects of health payments.(65;66)

3. New Analysis

a) Data sources

Considerable effort since the publication of the WHR2000 has been devoted to interacting with countries on possible sources of data. We have now analysed 98 income and expenditure surveys from 74 countries and continue to work with Member States to identify sources of data that can be analysed jointly.

b) Subsistence Expenditure

In response to the criticism that food expenditure is not a good indicator of subsistence expenditure, a method based on the international poverty line has been developed. (67) To translate the 1\$ per day (in international dollars) in 1985 prices into local currency units, we inflated to current period values using the US GDP deflator (the deflator for international dollars), then converted into local currency units using a food purchasing power parity for each country. Finally, household size was used to adjust estimates from an individual to the household level.

Household capacity to pay (CTP) is defined as total expenditure net of the expenditure implied by the international poverty line as long as total expenditure exceeds the poverty line.

In cases where total expenditure is less than the poverty line, CTP is taken to be the actual non-food expenditure.

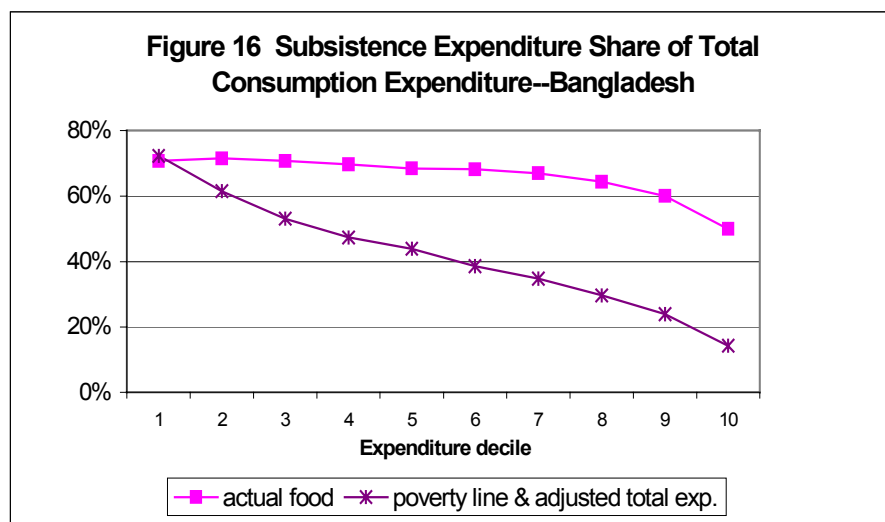


Figure 16 depicts the share of total consumption expenditure contributed by subsistence expenditure using the two definitions of subsistence expenditure, for Bangladesh. The vertical axis is the proportion of total expenditure and the horizontal axis divides households by expenditure decile. The first decile includes the households with the lowest 10% of reported expenditure and decile 10, those with the highest 10% of expenditures. When actual food expenditures are used as subsistence expenditure (the top line), richer households devote a lower proportion of their total expenditure to subsistence needs. But this decline with increases in expenditure is much more dramatic if subsistence expenditure is defined in terms of the international poverty line as outlined above.

This revision of the definition of subsistence expenditure raises the capacity-to-pay of richer households. This implies that the point where all households make an equal sacrifice, that is equal shares of capacity-to-pay, is more progressive. Capacity to pay is total household expenditure minus subsistence expenditure. Because subsistence expenditure decreases with rising total expenditure, households must contribute an increasing proportion of total expenditure to the health system to maintain an equal share of CTP.

c) Summarizing the distribution of HFC

In response to discussion about the index, we have compared different distributions of HFC using various summary measures. (68) Table 8 shows the results for 60 countries using for the Theil index, Mean Log Deviation (MLD), the Atkinson's index with deferent epsilons and the standard deviation. It is notable that they all give quite similar rankings across countries. Also notable is that the fraction of households facing catastrophic health payments (PHC) – defined to be when a household pays more than 40% if its capacity-to-pay for health – is highly correlated with all of the summary measures of the distribution of HFC.

Table 8. The Rank Correlation Coefficient of Different Inequality Measures

	Based on mean HFC						Based on sum mean HFC					
	FFCm	FFCom	Std(m)	MLD	Theil	A1	A2	A3	A4	FFCs	FFCos	Std(s)
FFCm	1											
FFCom	1	1										
Std(m)	0.9662	0.9662	1									
MLD	0.9537	0.9537	0.975	1								
Theil	0.9604	0.9604	0.9975	0.9732	1							
A ₁	0.9653	0.9653	0.9977	0.9748	0.9992	1						
A ₂	0.9680	0.968	0.9977	0.9753	0.9987	0.9998	1					
A ₃	0.9698	0.9698	0.9978	0.9753	0.9983	0.9994	0.9998	1				
A ₄	0.9702	0.9702	0.9976	0.9753	0.9981	0.9992	0.9996	0.9999	1			
FFCs	0.9795	0.9795	0.9628	0.9552	0.9581	0.9628	0.9652	0.9657	0.9659	1		
FFCos	0.9795	0.9795	0.9628	0.9552	0.9581	0.9628	0.9652	0.9657	0.9659	1	1	
Std(s)	0.9153	0.9153	0.9688	0.9492	0.9679	0.9666	0.9655	0.9649	0.9639	0.951	0.951	1
PHC	0.9602	0.9602	0.9557	0.949	0.9554	0.9585	0.9599	0.9596	0.9602	0.957	0.957	0.9176
A ₁	Atkinson index (e=0.20)				FFC	New FFC summary measure						
A ₂	Atkinson index (e=0.25)				FFCo	The one used in WHR200						
A ₃	Atkinson index (e=0.30)				(m)	Mean HFC was used in the measure						
A ₄	Atkinson index (e=0.35)				(s)	HFC norm (HMCo) was used in the measure						

Given the concern to have a summary measure that is more likely to have interval scaling properties, it is advantageous to convert the summary measure back into natural units. In other words, if we were to continue taking the cube of each household compared to the mean, we would want to take the cube root of the resulting aggregate. Moreover, following the logic of equal sacrifice, each household should in fact be compared not to the mean HFC but to the HFC that if applied to all households would generate the same amount of health system revenue. More formally,

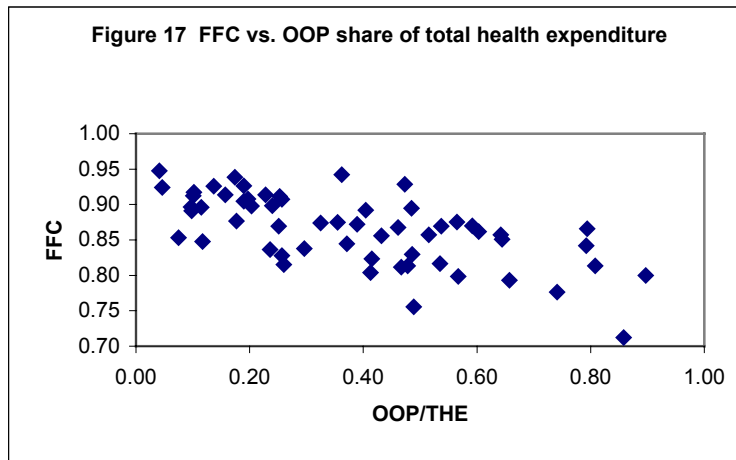
$$HFC_o = \frac{\sum HE_i}{\sum CTP_i}$$

A measure that fulfills these characteristics would include:

$$FFC = 1 - \sqrt[3]{\frac{\sum_{i=1}^n |HFC_i - HFC_o|^3}{n}}$$

This index still belongs to individual-mean family and it is an absolute measure of inequality.(62) It retains the property of the earlier FFC of putting a larger weight on households with catastrophic expenditure. The cube root function makes the new index more interpretable.

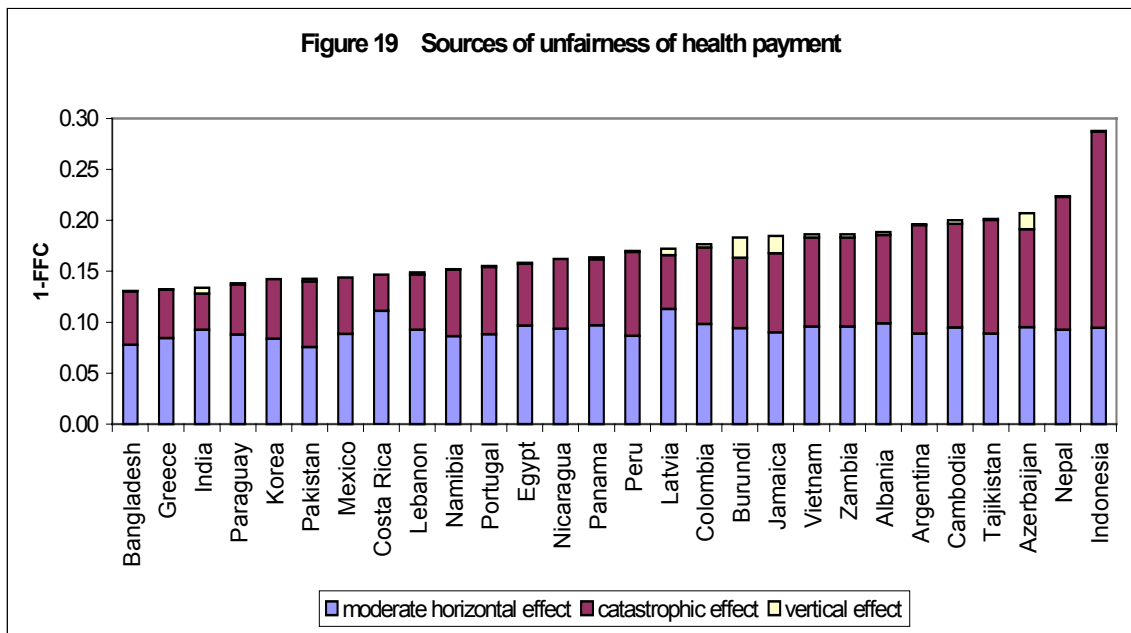
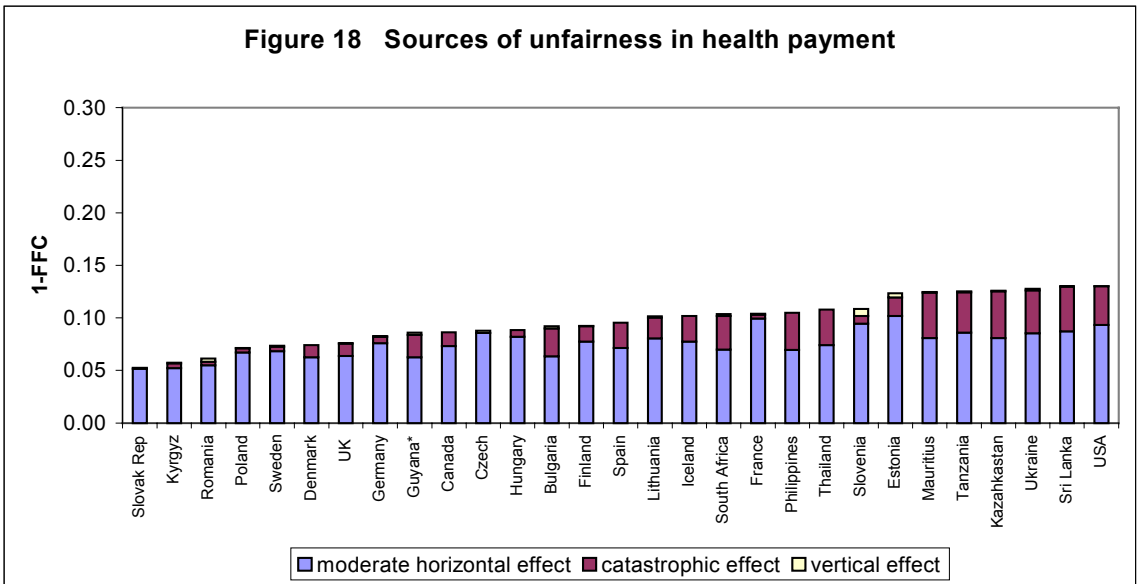
This version of the FFC is strongly correlated with out-of-pocket expenditure (OOPs) across countries (Figure 17) – in general, the higher is OOPs, the more unequal are household financial contributions to the health system. There is, however, considerable variability in the FFC for any given level of OOPs, so some systems with high OOPs still seem to be able to assure relatively equal household financial contributions.



d) Decomposition of the FFC index Summarizing the distribution of HFC

The FFC can be decomposed into different components – the parts due to extreme horizontal inequality related to catastrophic health expenditure, mild horizontal inequality and vertical inequality. This was done firstly by removing the effect of progressivity and regressivity in payments, then the effect of catastrophic expenditure recalculating the FFC each time. Some inequality remained, termed mild horizontal inequality (for full details, see the background documents). (69)

For countries with a lower value of the FFC index (or a higher value of 1-FFC in figures 18 and 19 - i.e. more inequality in household financial contributions), this analysis suggests that inequality is primarily attributable to household catastrophic spending. For countries with a higher FFC value (and a lower 1-FFC), moderate horizontal inequality is the dominant effect because there are only a few households with catastrophic health spending. The effect of vertical inequality is quite small. Although it could be hypothesized from theory that countries with very progressive tax systems may have a lower FFC index, the empirical results do not support this contention.



e) The impact of health system contributions in the space of income and the space of burden

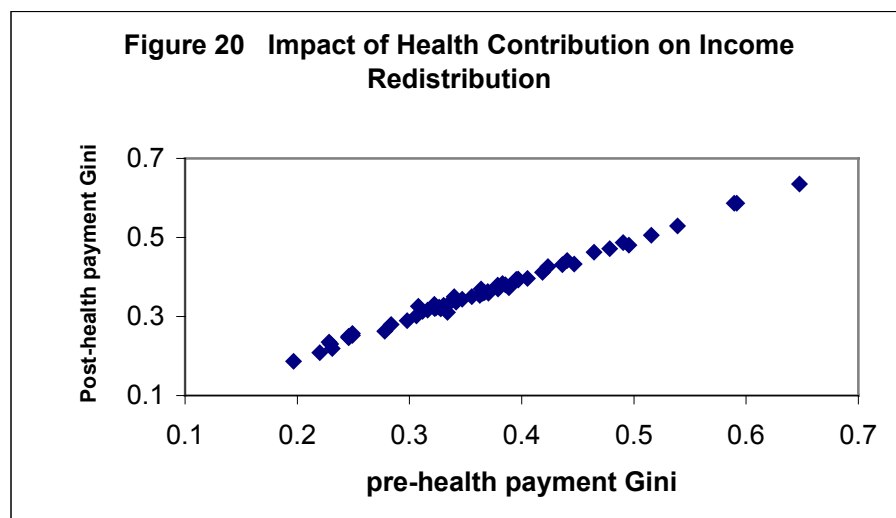
It is clear from the consultations that there is a constituency which values information about the marginal income redistributive effects of financial contributions to the health system. This concern is fundamentally different from a concern about deviations from equal sacrifice captured by the distribution of health financing contributions. It is, perhaps, useful to think of two distinct spaces of interest- the first is the space of income and the second is the space of

burden or sacrifice. Table 9 shows that it is possible to define indicators to capture the effects of health system contributions in both of these spaces, some of which summarize the continuous distribution of payments, and others based on thresholds used to divide the continuous distribution into different categories.(70)

Table 9 Approaches to the Distribution of Financial Contributions

Space	Distribution	Threshold
Income	Change of income distribution Redistributive effect (RE)	The difference in percentage of households below poverty line before and after health payment (DH)
Burden	Distribution of the burden FFC index	Percentage of households facing catastrophic spending (PHC)

Building on a long tradition of public finance analysis, it is possible to examine the marginal effect of health system contributions on the distribution of income (66). In most systems, the impact of the health system on the distribution of income as assessed by the Gini coefficient is rather small (figure 20).



Nevertheless, the extent to which the health system at the margin may increase or decrease income inequality may be an important policy issue. It is interesting to note that while most of the literature has quantified the impact on income distribution using metrics derived from the Gini coefficient, other approaches to quantifying changes in income distribution could also be used.

Recently Wagstaff and van Doorslaer have proposed to extend a concern about the effects of health system contribution in the space of income to examine the number of individuals that are pushed into poverty. (66) Using household data from Vietnam, they used several alternative measures of impoverishment including headcount measures and depth of poverty measures to study the impacts. To explore this phenomenon more extensively, we have analyzed impoverishment in 57 countries using the international absolute poverty line. (70)

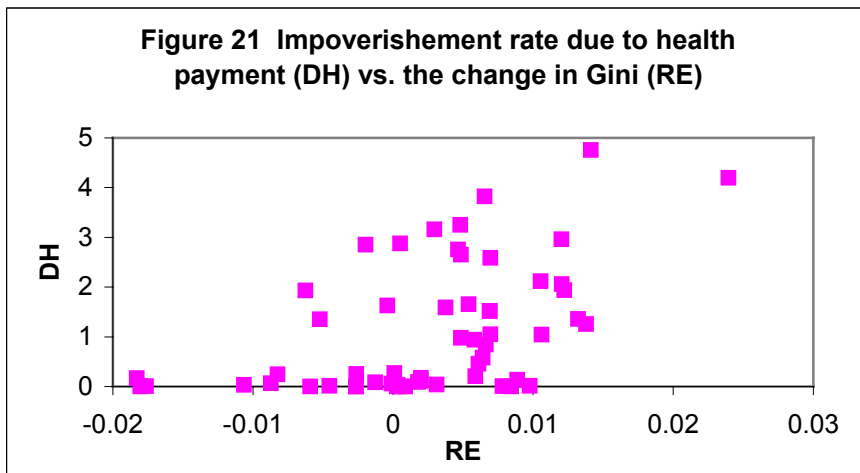


Figure 21 shows a scatter plot of impoverishment rates and changes in the Gini (also known as the redistributive effect or RE). Examining the marginal effects of health system contributions on the distribution of income or the number of individuals in poverty gives startlingly different results across countries. This strengthens the Wagstaff and van Doorslaer (66)proposition that these types of analysis provide alternative perspectives.

We can also examine the impacts of health system contributions in the space of burden or sacrifice imposed on households. A concern about equalizing the sacrifice imposed on households through health system contributions is not a concern about the marginal effect of health system contributions in the space of income. In the space of burden of sacrifice, there can be both an interest in the distribution of sacrifice and those that face sacrifice above some critical threshold, namely catastrophic payments. The distribution of HFC or burden reflects both catastrophic payments and less severe forms of unequal sacrifice. Table 10 provides estimates of the distribution of HFC and catastrophic payments (PHC) for 57 countries.

Table 10 The FFC index and the percentage of households with catastrophic expenditure (PHC)

country	PHC	FFC	country	PHC	FFC
Albania	5.72	0.86	Lithuania	0.69	0.90
Argentina	8.10	0.82	Mauritius	0.95	0.89
Azerbaijan	4.86	0.85	Mexico	1.98	0.86
Bangladesh	1.73	0.89	Namibia	2.92	0.86
Bulgaria	0.20	0.91	Nepal	6.82	0.79
Burundi	3.41	0.82	Nicaragua	3.89	0.86
Cambodia	5.81	0.83	Pakistan	1.45	0.88
Canada	0.48	0.92	Panama	2.44	0.86
Colombia	5.50	0.85	Paraguay	2.01	0.88
Costa Rica	2.30	0.87	Peru	2.82	0.85
Czech	0.00	0.91	Philippines	0.71	0.92
Denmark	0.28	0.93	Poland	0.18	0.93
Egypt	2.95	0.86	Portugal	4.80	0.84
Estonia	2.23	0.89	Romania	0.04	0.94
Finland	0.78	0.91	Russia	8.55	0.79
France	0.37	0.90	Slovak Rep	0.00	0.95
Germany	0.37	0.92	Slovenia	1.27	0.90
Greece	2.44	0.87	South Africa	0.67	0.90
Guyana	0.60	0.93	Spain	0.71	0.91
Hungary	0.35	0.92	Sri Lanka	1.70	0.89
Iceland	0.83	0.90	Sweden	0.39	0.93
India	1.90	0.88	Tajikistan	6.45	0.81
Indonesia	7.76	0.74	Tanzania	1.75	0.89
Jamaica	3.83	0.83	Thailand	0.87	0.90
Kazakhstan	1.15	0.89	UK	0.24	0.93
Korea	2.17	0.86	Ukraine	1.06	0.88
Kyrgyz	0.05	0.94	USA	2.45	0.87
Latvia	4.97	0.84	Zambia	5.00	0.81
Lebanon	7.67	0.86			

We have also compared the correlation of the measures of threshold and distribution indicators, reported in Table 11. The redistributive effect or change in the Gini (RE), a commonly used summary of the impact of payments on income redistribution, is not highly correlated with the indicators of inequality in burden, nor is it correlated with the threshold

indicator in income space. On the other hand, in the burden space the two types of measures – the FFC and the proportion of people with catastrophic expenditures – are highly rank correlated. One of the policy implications of this is that the RE and the FFC measure different, but complementary, concerns so that it might well be worth measuring both.

Table 11 Correlation coefficients: Indicators in Income and Burden Space

	FFC	PHC	RE	DH
FFC	1.000			
PHC	-0.870	1.000		
RE	-0.143	0.182	1.000	
DH	-0.685	0.647	0.356	1.000

f) Case study on catastrophic health expenditure

Catastrophic health expenditure is highly correlated with the FFC across countries. (71) In fact, we showed above that the unfairness of the health financing systems where FFC is very low is mostly attributable to the higher percentage of households facing catastrophic expenditure. For policy-makers it is important to know which households face higher risks of catastrophic health expenditure.

We illustrate using Lebanon as an example.(72) Table 12 reports a regression in which having catastrophic expenditure is the dependent variable. The dependent variable is categorical, taking the value of 1 for households with categorical expenditures and 0 otherwise.⁸ The existence of chronic disease, having someone with a handicap in the family, and the presence of old people in the household are risk factors increasing the probability of catastrophic health expenditure. On the other hand insurance coverage, higher income, larger family size, higher levels of education for the household head, and being native to Lebanon are protective factors.

Table 12 Results of Logistic Regression

Catastrophic	Coef.	Std. Err.	z	P>z	Odds Ratio
Chronic disease	0.362	0.045	8.10	0.000	1.436
Elderly members	0.401	0.072	5.53	0.000	1.493
Insurance	-0.642	0.125	-5.16	0.000	0.526
Handicapped	0.372	0.120	3.10	0.002	1.450
Beqaa	0.611	0.157	3.89	0.000	1.842
Household size	-0.102	0.027	-3.75	0.000	0.903
Lebanese	-0.478	0.219	-2.19	0.029	0.620
Primary & secondary school dummy	-0.303	0.127	-2.39	0.017	0.739
University degree dummy	-1.318	0.429	-3.07	0.002	0.268
per capita income	-0.001	0.000	-3.18	0.001	0.999
Nabatieh	0.372	0.172	2.16	0.031	1.451
Constant	-2.377	0.260	-9.15	0.000	

⁸ Catastrophic expenditure defined as greater than 40% of capacity to pay.

From the estimated equation it is possible to predict a household's risk of facing catastrophic expenditure. Take two households. The first family lives in Beyrouth (e.g. not in Nabatieh), has per capita income of 500,000 Lebanese pounds, three members one of whom is covered by health insurance, the household head has a university education and is Lebanese. The second household has a per capita income of 100,000 Lebanese pounds, contains five members, no-one is covered by health insurance, two members have chronic diseases, one is handicapped, the household head has no formal schooling and is not Lebanese. The logistic regression suggests that the risk of facing catastrophic health expenditure is only 0.31% for the first family while is 21% for the second family.

4. Proposals – Fairness of Financial Contributions

WHO proposes:

- 1. All four types of indicators in Table 9 should be monitored and reported. In this way, information on the impact of health system contributions on income distribution, poverty rates, the distribution of sacrifice and catastrophic payment will be available to policy-makers.**
- 2. Household capacity to pay should be based on the revised definition where subsistence expenditure is derived from the international absolute poverty line.**
- 3. The distribution of HFC should be summarized using the root mean cube difference between households and the equal sacrifice HFC.**
- 4. Catastrophic health spending should be defined as spending more than 40% of a household's capacity to pay to the health system.**
- 5. In collaboration with Member States, low cost information generation strategies should be developed for obtaining data on financial contributions where they are not available. This should be linked to the World Health Survey.**

D. Responsiveness

8. The WHR2000

Responsiveness was included as an intrinsic goal because the way people are treated when they come in contact with the system can improve or reduce well-being independent of health outcomes. Although the literature on quality assurance and patient satisfaction provides useful insights on the aspects of population-system interactions that were important to people, responsiveness is concerned with what actually happens to people when they come in contact with the system rather than their expectations or opinions about their experiences. It was defined in the WHR 2000 to include only the non-health related components of the interaction with the system because the health related components were captured by the intrinsic goal of improving population health.

The initial list of domains of responsiveness that were field tested, then measured for the WHR2000 included dignity, autonomy, confidentiality, and prompt attention (collectively called respect for persons) and access to social support, quality of basic amenities and choice (called client orientation). Measurement was based on a key informant survey in 35 countries, from all 6 WHO Regions. To estimate responsiveness for the remaining countries, a similar approach to the one used for fair financing and health inequalities was used. The results for the 35 countries were regressed on a set of covariates, and the relationship used to estimate responsiveness for countries without surveys.

9. Commentaries on the WHR2000

Most participants in the consultations agreed that people wanted their systems to be responsive to their legitimate non-health expectations, although some thought that it was also important to measure patient satisfaction.(73;74) While the seven domains of responsiveness identified in the WHR2000 were generally considered to be relevant, it was argued that other domains might also be important. The technical consultation on this topic recommended that an additional domain should be added focusing on communication between providers and the population. The other suggestion of a conceptual nature was that all societies would not value the different domains in the same way, a testable hypothesis.

Other commentators focused on the way that responsiveness was measured in the WHR2000. The use of key informant surveys to measure both the level and the distribution was criticised on the grounds that key informants would not necessarily reflect the experiences of the general population with their health systems. And even if they did, it would be impossible to estimate inequality in responsiveness in a valid, reliable and comparable way from expert opinion. A final concern was that the instrument used in the key informant surveys in the WHR2000 focused solely on personal health services, which was not consistent with the definition of health systems which included personal, non-personal and inter-sectoral health actions.

10. New Analysis

a) Satisfaction versus Responsiveness.

We explored some of the results from published satisfaction surveys in OECD countries. Among their findings is that satisfaction with the way the health system is run is higher among the poor than the rich in 9 of the 17 studies for which survey data were available.(74) Yet we strongly suspect that the poor are not treated better than the rich. While satisfaction may be important as an indication of public sentiment, we argue that the interest of an organization such as WHO should lie with the way people are actually treated when they come in contact with the system. This approach was supported by the technical consultation on the topic, although not by all participants in the regional consultations.

b) The Domains of Responsiveness.

Questions relating to the new domain of communication were developed in response to the consultations and included in the WHO Multi-country Survey Study of 2000-2001.(75) Experience with this domain and the confirmatory factor analysis outlined below support the addition of this domain.

c) Survey Mode

In response to the consultation process, the gold standard method for deriving information on responsiveness was taken to be population face-to-face interview surveys. In some settings, Member States have expressed a desire for other lower cost appraisal methods. In order to establish the validity of alternative information collection modes, the WHO Multi-country Survey Study undertaken in 2000-2001 included different types of representative population samples, including brief face-to-face interviews, telephone and postal surveys. An overlap design was used whereby two types of surveys were undertaken in each of 10 countries, allowing the results of the different modes to be compared. A total of 71 household surveys were undertaken in 61 countries – full details are found in section VII. At the same time, 28 additional key informant surveys have been undertaken (involving 10,500 respondents) in an

attempt to test if they provide valid estimates of the average level of responsiveness compared to the gold standard.

It is possible that a combination of survey modes might be shown to be useful. For example, key informant interviews might provide better information on confidentiality of information, for example, than household surveys where few people are aware of their rights and the system's practices regarding medical records. Legal experts would have a detailed knowledge of this aspect of responsiveness.

d) Cross-Population Comparability

The problem of cross-population comparability described earlier in relation to measuring health applies to responsiveness as well. The survey instrument asked respondents to rate their last contact with the health system on the different domains of responsiveness. They were offered five possible response categories – very good, good, moderate, bad and very bad. For any given level of autonomy, for example, people categorize their experiences in different ways, implying that their cut-points between the possible categorical responses differ (as explained earlier with respect to self-reported health).

To establish how individuals use the categorical responses very good to very bad, a series of vignettes were devised for each domain covering the entire range of the latent or unobserved variable. For each vignette, respondents were asked to rate the experience described as very good to very bad. The responses on the set of vignettes for each domain have been analyzed using the hierarchical ordered probit model to identify how cut-points on the latent scale systematically vary across individuals and across communities. This information allows a more meaningful interpretation of each individual's responses for their own encounter with the health system. More detail is provided in section VII and the background documentation.⁽⁷⁶⁾ Some of the vignettes used in the Survey Study did not perform well in that they did not elicit a response that was significantly different to other vignettes. Now that it has been shown that the vignettes are useful ways to control for different cut-points, it remains to finalize a set of vignettes for each domain that cover the full range of the latent variable of interest.

e) The Survey Instrument

Traditional psychometric tests were used to analyse the strengths and weaknesses of the questionnaire used in the WHO Multi-country Survey Study. Item missing rates, factor analysis and the calculation of Kappa values were some of the classical techniques used to examine the instrument's validity and reliability.⁽⁷⁷⁾ Exploratory and confirmatory factor analysis showed that the items loaded on the domain factors with which they were identified. Figure 22 below shows the factor loading results for confirmatory analysis on 67 surveys. The question numbers and the factor loadings associated with each of the items are shown for each domain. The factor loadings in most cases, except for some of the questions associated with outpatient prompt attention, were higher than 0.70. This means that respondents associated the questions asked in the survey with the underlying eight responsiveness domains.

Outpatient							
Item	Prompt Attention	Dignity	Communication	Autonomy	Confidentiality	Choice	Quality of Basic Amenities
q6100	0.302						
q6101	0.636						
q6103	0.336						
q6104	0.922						
q6110		0.922					
q6111		0.884					
q6112		0.786					
q6113		0.814					
q6120			0.869				
q6121			0.903				
q6122			0.891				
q6123			0.828				
q6131				0.841			
q6132				0.825			
q6133				0.842			
q6140					0.806		
q6141					0.954		
q6142					0.786		
q6150						0.895	
q6151						0.866	
q6152						0.753	
q6160							0.905
q6161							0.908
q6162							0.943
Inpatient*							
Item	Prompt Attention	Social Support					
q6302	0.745						
q6303	0.892						
q6304	0.897						
q6311		0.747					
q6312		0.826					
q6313		0.792					

* Presented here only include two domains, prompt attention and social support. For the other domains of Inpatient section, each domain has only one item

Figure 22: Confirmatory Factor Analysis: factor loadings

Three other issues are important for the survey instrument. First, the current instrument focuses on the personal interactions between providers and members of the population. We are currently developing questions on non-personal services but these are more complex to communicate to people and to standardize across settings. One example of such a question would be how people view the clarity of communication around awareness campaigns for lung cancer, breast cancer or HIV.

Second, earlier instruments used by WHO have focused only on people who have used personal health services. The consultations suggested that questions around the non-use of services should be piloted, in particular to explore if perceived lack of responsiveness causes people to avoid the health system or part of it – e.g. the public sector.

Third, the Survey Study excluded parents reporting for children so there was no evidence about responsiveness for anyone under 18 years old. The consultations suggested that we ask parents about their children’s interactions with the health system up until the age of 12. While this approach still omits the remaining 13-17 year olds, it will increase coverage of the population. While desirable for completeness, interviewing the age group 13- 17 years old would be operationally difficult in many countries as experts do not recommend proxy reporting for children older than 12 years. We are currently testing ways of asking parents about their encounters with younger children.

f) Ranking of Domains

In the survey people were asked to identify which domain or domains of responsiveness were most important to them and which domain or domains were least important. Using the ordered probit model and the requirement that weights on the eight domains must sum to one and the least important domain is bounded by zero and 0.125, it is possible to empirically assess variation in the implied weights across countries. Figure 23 shows the range of the

implied weight for each of the eight domains from this analysis .(79) The results confirm that the ordering and weighting of domains of responsiveness are similar across countries. In 45 of 50 countries, prompt attention received the highest weight and in 47 of 50 countries social support received the lowest weight.

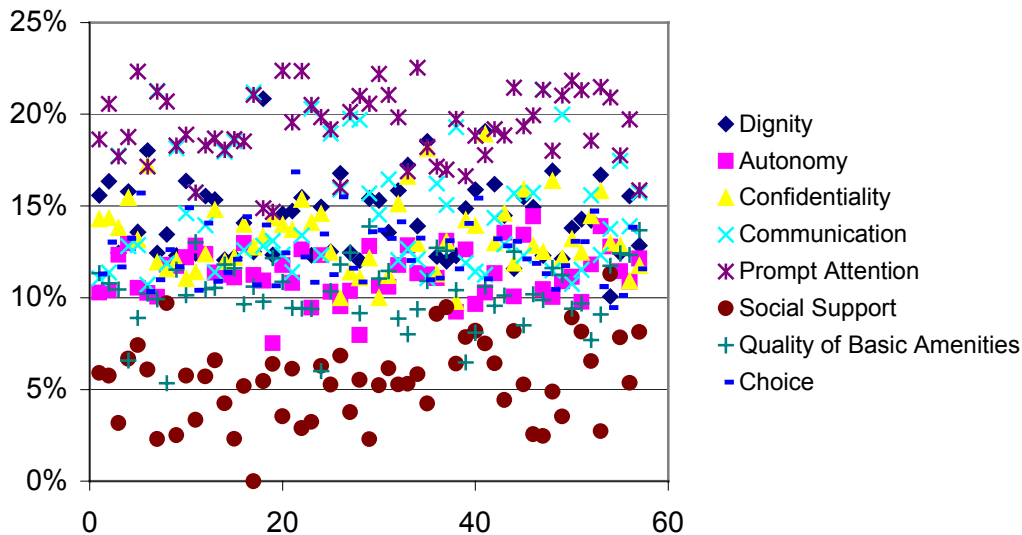


Figure 23: Comparison of Domain Weights

g) Low inpatient encounter rates

The WHO Multi-country Survey Study suggested that representative population studies might not provide sufficient numbers of people with severe conditions to understand what happens to them when they have an inpatient experience. The current questionnaire restricts reporting on outpatient and inpatient visits to the previous 12 months. For every household survey of 1000 people, an average of 500 may report outpatient encounters and as little as 100 report inpatient encounters in the previous 12 months. Technical experts had suggested that this time period was the maximum possible to have a reasonable recall of all possible interactions with the system.

There are two strategies available to try and increase the number of inpatient experiences captured in the household survey sample. First, the recall window for inpatient services could be extended on the grounds that people remember this type of intense encounter for longer periods. The consultation suggested the recall window for inpatient services be extended to two years. The interview would also document the time of the last inpatient episode to enable assessment of any bias created through longer recall periods. Second, the within-household sampling strategy could be modified so that any household member who has had an inpatient experience would be interviewed. This would effectively increase the number of inpatient events by the average size of the household.

h) Country Responsiveness

The Multi-country Survey Study allowed estimates of country level responsiveness to be computed based on the representative population surveys. Responsiveness across the countries ranged from a minimum of just over 35 in Greece (maximum possible is 100) to 95 in the Canadian telephone survey. Although it is positively correlated with health expenditure per capita across countries Figure 24 there are large variations around the trend. For example, at an expenditure level of approximately I\$1000 per capita, there is a three-fold difference in responsiveness across countries.

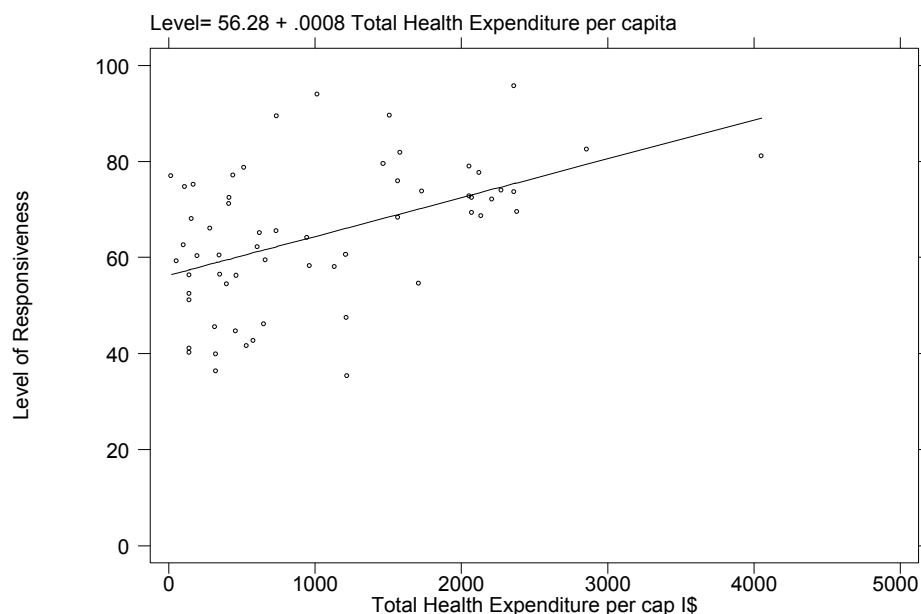
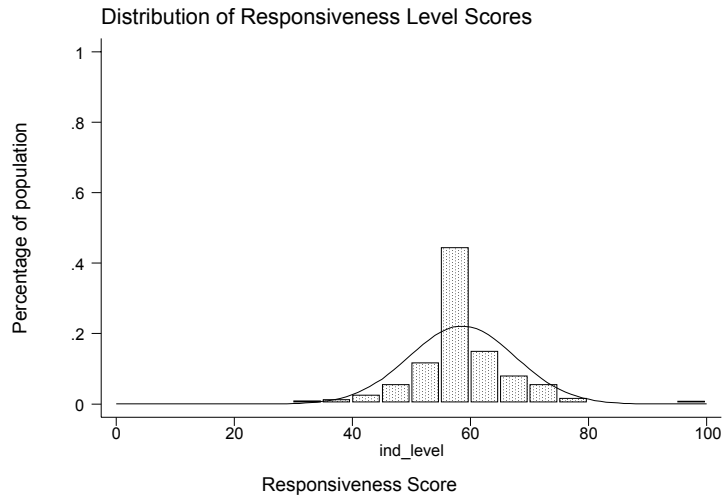


Figure 24: Country Mean Level of Responsiveness and Total Health Expenditure per Capita

i) Inequality in Responsiveness

Household survey data on responsiveness provides the opportunity to evaluate the distribution of responsiveness across the sample of individuals. (80) If the sample frame is representative and properly executed, the resulting distribution of responsiveness should provide a strong basis for evaluating the distribution of responsiveness in the population. To illustrate, Figure 25 reports the distribution of responsiveness across the respondents to the Multi-country Survey Study 2000-2001 from Spain and Morocco. The average level of responsiveness is higher in Spain than in Morocco, and there also seems to be less variation in the way the system responds to people in Spain.

Spain



Morocco

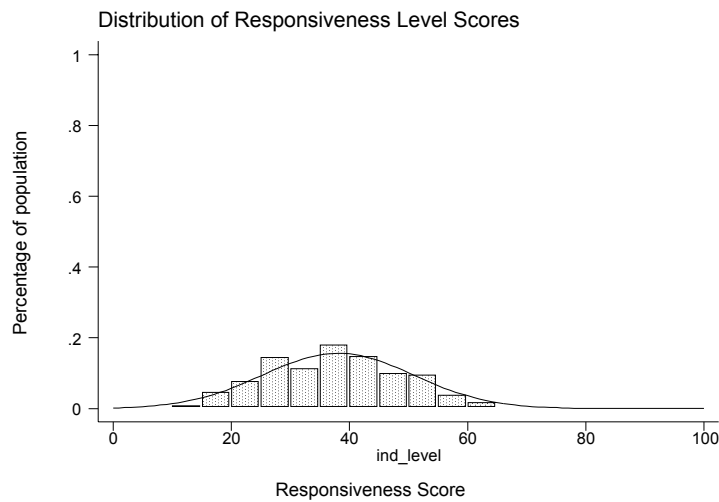


Figure 25: The Distribution of Responsiveness in Spain and Morocco

11. Proposals – Responsiveness

WHO proposes to:

1. Continue improving the household survey instrument on responsiveness by improving the vignettes that assure cross-population comparability, dropping poorly performing items, and including questions focusing on non-personal health services and the interactions of children with the system. A further round of testing will take place in the pre-pilot phase of the World Health Survey.
2. Include a module of responsiveness in the World Health Survey;
3. Measure the distribution of responsiveness (inequality) using household survey data.
4. Continue testing ways of supplementing this information with other data collection strategies including exit interviews, facility surveys, and key informant interviews.

E. Composite Goal Attainment

1. WHR 2000

A composite index of goal attainment was constructed for each Member State in the WHR2000 as a weighted sum of attainment on the individual intrinsic goals. Weights were obtained from a world-wide-web key informant survey involving over 1600 participants, largely public health practitioners, from over 100 countries. Fifty percent of the total weight was ascribed to health (25% to the level and 25% to inequalities), 25% to fairness of financial contributions, and 25% to responsiveness (12.5% level and 12.5% inequalities). Uncertainty intervals were reported for the scores on the attainment index and the associated ranks.

2. Commentaries on the WHR2000

A number of people at the regional consultations expressed the opinion that WHO should not seek to combine attainment on the individual goals but should simply report attainment on each goal separately. Other suggestions were:

- countries have different weights for the goals depending on their backgrounds and the stage of development; (81;82)
- diverse disagreements with the weights that were used ranged from the opinion that 100% should be given to health, the defining goal of the system; too much weight was given to equity; too little weight was given to equity; or that the relative weights would change according to the level of attainment on the individual goals;
- key informant interviews, particularly those based on health professionals with access to the world-wide-web, are not representative of population preferences; (63;64)
- the web-based questions on weights did not present the questions as trade-offs, but asked people to divide a pie into sections.

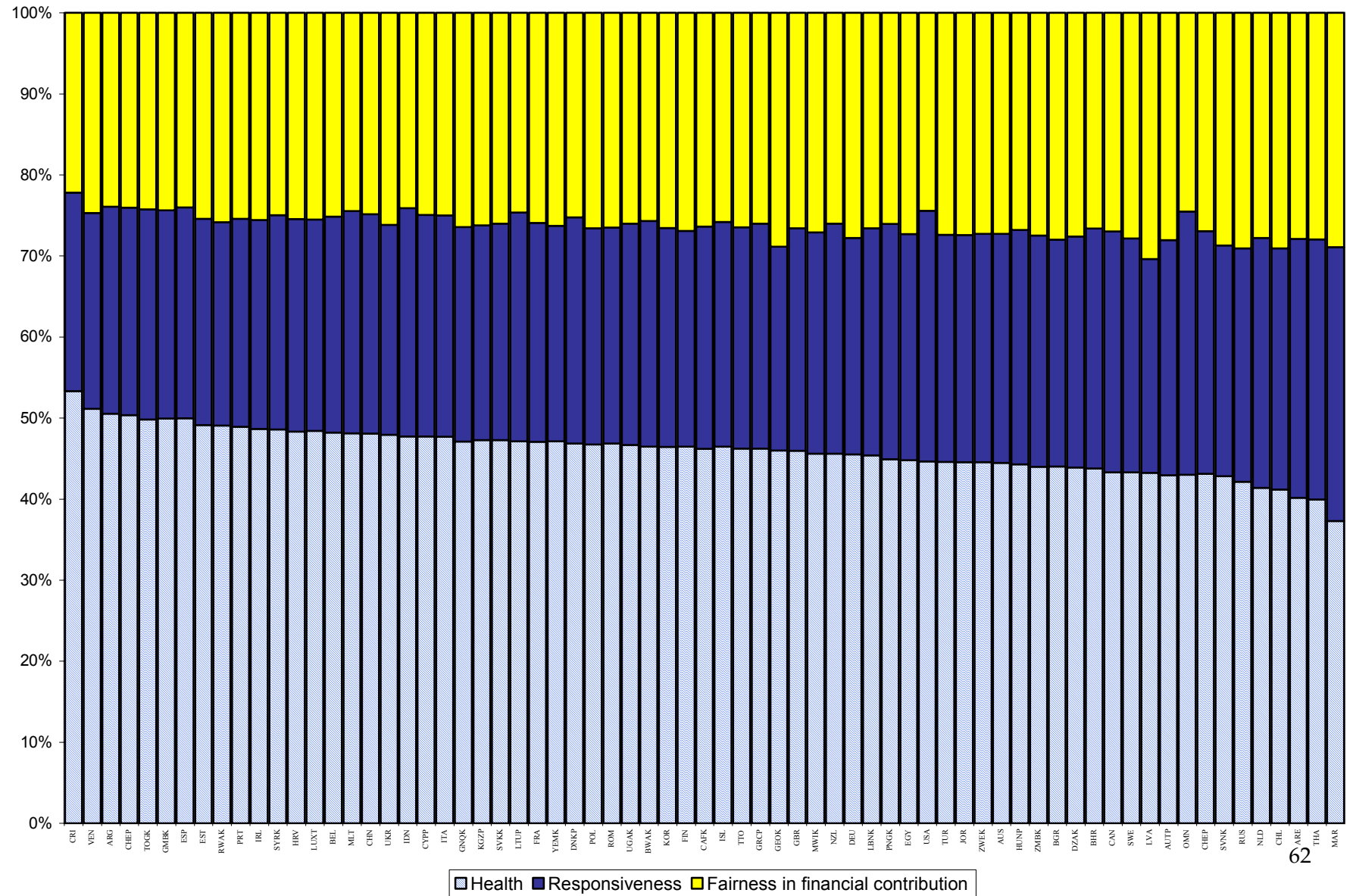
3. New Analysis

a) Weights

It is a testable hypothesis whether a web-based survey of health professionals does not produce the same weights as questions directed at a representative population sample. We are in the process of testing it with the results of the WHO Survey Study where we have analysed responses from over 51,000 respondents from representative households in 53 countries, and have an additional 10,500 key informant interviews in 28 countries to analyse.(75)

Weights as assessed through the questions in the Multi-country Household Survey Study vary across countries (Figure 26). The extent of variation, however, is perhaps smaller than believed by some. In no household survey were the average reported weights equal to zero for any component. Empirically, the claim that only health is important to populations is not supported by this body of survey data.(83) The average weights across those surveys that have been analysed are 46% for health (25% for level and 21% for distribution), 26% for fairness of financial contribution, and 28% for responsiveness (15% for level and 13% for distribution). Overall these weights are similar to the weights recorded in the internet survey conducted in 2000. Nevertheless, it is worth noting that on average the data from representative household surveys has given more importance to responsiveness and fairness in financial contribution. Analysis of within country systematic variation in weights has revealed a number of statistically significant relationships but none of these relationships are substantively large.

Figure 26. Weights given to Health , Responsiveness and Fairness of Financial Contributions, WHO Multi-Country Survey Study 2000-2001



b) Benefit-of-the-Doubt Scores.

We have recalculated the overall attainment index reported in the WHR2000 to take into account variations in the weights observed in the Survey Study.(84) Each country's weights were allowed to vary between the minimum and maximum scores observed across survey countries. For each individual country, weights were chosen from within this range so as to maximize the overall attainment of the country. Given the individual component scores, this procedure results in the highest possible overall attainment score for each country and is therefore termed the "benefit-of-the-doubt" score. Results are reported in Figure 27.

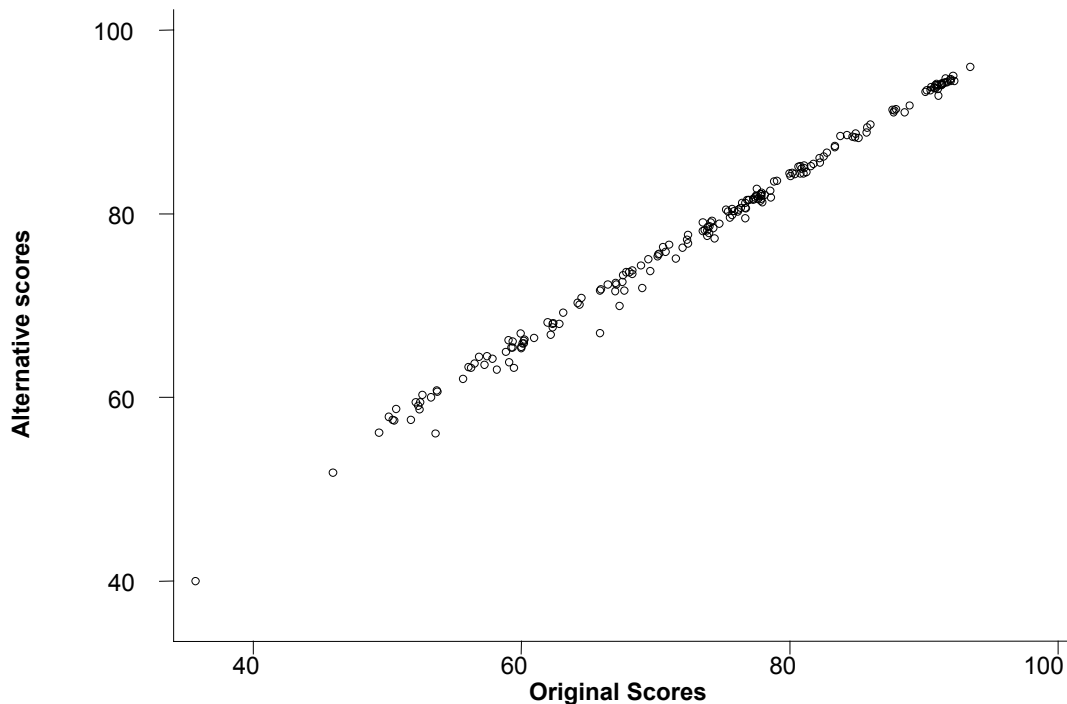


Figure 27. Overall Attainment using Benefit-of-the-Doubt Weights vs. WHR2000 weights

As expected, the benefit-of-the-doubt attainment scores are consistently higher than the scores reported in the WHR2000, but the correlation between the two sets of scores is, in fact, very high (0.9978). In addition, while rankings of countries changed a little as a result of using benefit-of-the-doubt weights, all ranks remained within the uncertainty intervals reported in the WHR2000. The simple correlation coefficient of country ranks between benefit-of-the-doubt scores and WHR2000 scores was 0.9972.

4. Proposals - Composite Attainment Index

WHO proposes that:

1. The composite index should be calculated and reported routinely. Those who prefer to focus on components instead can still do so since the individual component scores would also be reported. In this way, the composite measure of goal attainment would be available for those users who are interested in it.

2. To provide some basis for comparability, the average weight across countries should be used to estimate the index of overall social goal attainment;
3. In addition, overall attainment using benefit of the doubt weights should be reported, as should an index based on the weights reported for each country.
4. WHO should continue to investigate whether there are systematic determinants of country weights, including the level of attainment on particular goals. Alternative forms of eliciting weights for the goals including tradeoff questions should be explored.

VI. EFFICIENCY

A. WHR2000

Referring back to Figure 1, the WHR2000 estimated efficiency as $e/(e+f)$. It was the ratio of attainment (above the minimum) to the maximum possible attainment (also above the minimum) – i.e. what proportion of the potential health system contribution to goal attainment is actually achieved for the observed level of resources.

The minimum possible level of health that would be observed in the absence of the system was estimated from the relationship observed between literacy and life expectancy at the turn of the century, a point at which the health system made little contribution to population health. This relationship was used to estimate the current level of health that would have been observed with current literacy levels if the system had not existed. The maximum level, and the associated efficiency estimate for each country, were derived using frontier production function estimation techniques described elsewhere.⁽⁸⁵⁾

A fixed effects model with panel data from 1993-1997 was estimated using a translog specification, the most general form of the production function. From this, the maximum possible outcome for observed levels of inputs can be estimated. Health expenditure per capita was used as the health system input and the average years of schooling of the adult population as the non-health system input to health. Accordingly, inefficiency could have been due to two causes. The first is waste – e.g. over-staffing of hospitals - or **technical inefficiency**. The second is that the wrong mix of interventions had been chosen for the health problems of the country and the available resources. This is sometimes called **allocative efficiency**.⁹

Efficiency was estimated in terms of health alone, the defining goal of the health system, and in terms of the composite attainment index. For the latter, the following assumptions about the minimum were made in addition to the minimum level of health that would be observed in the absence of the system. Inequalities in responsiveness and fairness of financial contributions are given full scores in the absence of a system. This is simply because a non-existent system treats everyone totally equally – it is totally unresponsive to all people (so there is full equality). It was assumed that the score for the other two components (health inequalities and responsiveness level) would be zero in the absence of a health system - a non-existent health system is completely unresponsive. Although health inequalities surely would exist in the absence of a health system, with respect to the health system goal of reducing inequalities, zero progress can be claimed, so a zero weight was used for this component.

⁹ Allocative efficiency the mix of inputs that minimizes cost for a given level of output. If interventions are viewed as inputs, allocative efficiency can be considered as the mix of interventions that is the most cost-effective. This is how the term is often used in the health economics literature.

B. Commentaries on the WHR2000

It was agreed at the consultations that the concept of efficiency adds another dimension to health system performance assessment. The Human Development Index of UNDP, for example, only measures attainment and the richest countries not surprisingly attain more than poorer countries. It is possible for poorer countries to use their resources efficiently even if they attain less than richer countries, and this should be taken into account when assessing performance.

Despite agreement in principle, considerable debate has been directed at a number of details. (86) These include:

1. *Timing.*

There are significant lags between the timing of health system inputs and system outcomes. Some of this year's health attainment is due to actions taken in the past, perhaps decades ago. Some of this year's health inputs will not improve health for some time into the future. These lags were not well accounted for in the WHR2000.

2. *The Minimum.*

Other determinants of the minimum exist and should have been included, including geographical location and epidemiology.

3. *Difficulty.*

The difficulty of translating inputs into outcomes was not fully captured in the production function. This would include variables such as the proportion of the population in isolated areas, for example.

4. *Determinants of Output.*

Two diverging viewpoints were expressed. On the one hand, additional variables to expenditure and education should have been included in the production function itself. (87) On the other hand, it was argued that if the minimum truly represented the outcome that would exist in the absence of the system – i.e. taking into account the current levels of non-health system inputs – the only variable that should be included in the production function is health expenditure. Because healthy life expectancy is a measure that is insensitive to age structure of the population, should inputs such as health expenditure be age-standardized? The composite index of goal attainment includes distributional measures of health and responsiveness. Logically, distributional measures of health expenditure and any other controllable factors of production should be included on the right-hand side.

5. *Determinants of Efficiency.*

Variables influencing efficiency had not been included in the estimation procedure. These are the variables of particular value to policy-makers, including the type of health financing system, or the mix between public and private provision of services.

C. New Analysis

1. *Determinants of Efficiency.*

Traditional frontier production function analysis generally separates the estimation of the production function from the explanation of efficiency. Variables in the first step are the inputs to the production process, and in the second are the policy variables which influence

efficiency. Sometimes the estimation of the two stages is undertaken in the same procedure, and sometimes as separate steps. We have analysed the determinants of efficiency as a second step.(88)

A number of possible explanatory variables were tested and the results (see Table 13) suggest that income inequality, the extent of the AIDS epidemic, and very low health expenditure per capita (less than \$60 in international dollars) were negatively correlated with health system efficiency in producing health. GDP per capita and a French legal system were positively correlated with this type of efficiency. These correlations were statistically significant at the 1% level. On the other hand variables such as population density, dependency ratios and the proportion of health expenditure contributed by the private sector did not prove to be significant.

	Coefficient	Std. Err	t	P> t
Gini (income inequality)	-3.24	0.92	-3.50	0.001
Govt. effectiveness	0.24	0.10	1.89	0.061
GDP per capita	0.000034	0.00001	2.52	0.013
Impact HIV/AIDS	-0.17	0.02	-7.20	0.000
Health expenditure < I\$60 per capita	-0.37	0.17	-2.70	0.008
Geographical location in Latin America	0.41	0.20	1.67	0.10
French legal system	0.43	0.16	2.75	0.007
French colonial background	-0.35	0.22	-0.90	0.40
Constant	2.04	0.35	5.82	0.000

Table 13 Multivariate analysis of determinants of efficiency on health¹⁰

When indicators of the functions described in section IV become available, it will be possible to include them as explanatory variables and they would have considerable policy relevance. At the same time, we are also working on developing a likelihood that enables the two stages of the process to be estimated at the same time, a more theoretically acceptable approach.

2. Random Coefficients Models.

Exploratory work is being undertaken in collaboration with acknowledged experts on frontier production functions to estimate a random coefficients model. This allows the impact of the inputs, or the impact of the determinants of efficiency, to be themselves modeled as functions of other variables.

There are two key aspects of this type of modeling strategy: (a) the determinants of output and the determinants of efficiency are modeled in the same likelihood function, and (b) the

¹⁰ $R^2 = 0.68$ and adjusted $R^2 = 0.66$

coefficients on the determinants of outputs and on determinants of efficiency are themselves modeled as being stochastic, i.e., the model is a random coefficients model.

For country i , let Y_i be output, X_i be a vector of inputs, and Z_i a vector of exogenous determinants of efficiency. Then the model decomposes the error such that:

$$Y_i = X_i \beta_i - u_i + v_i$$

where u_i is the (non-negative) technical inefficiency component of the error, v_i is random noise, and β_i is a country-specific coefficient reflecting the translation of inputs to outputs. In addition,

$$\beta_i = \beta + e_i$$

with e_i reflecting stochastic variation of the parameter across countries, and

$$u_i = \gamma_i Z_i + w_i$$

where w_i is random noise, and γ_i is a country-specific parameter reflecting the effect of exogenous factors on efficiency such that:

$$\gamma_i = \gamma + v_i$$

with v_i being random noise.

Recent methodological innovations allow for the estimation of such models using simulation-based techniques. One advantage of these models is that they are very general. Several simpler frontier models can be derived as restricted forms of these models. The other advantage is that these models allow for parameters relating to the determinants of output and determinants of efficiency to be different for different countries: this is more plausible than the assumption that model parameters are the same across all countries.

3. Multiple Indicator Models.

These models are based on the premise that health system output (Y) is a function of: (a) direct "inputs" or positive determinants (X), (b) the efficiency (E) of the health system in translating inputs into outcomes, and (c) other factors (D) which may have a negative effect on outputs, signifying the difficulty of the environment within which the health system functions:

$$Y = f(X, E, D).$$

The efficiency of the health system is not directly observed (it is a latent variable). What are observed are multiple indicators of this latent variable (E_1, E_2, E_3, \dots), each being related to the underlying efficiency E :

$$E_1 = f_1(E)$$

$$E_2 = f_2(E)$$

$$E_3 = f_3(E).$$

Indicator 1 can be considered to be macro estimates of production using cross-country panel data similar to the way that was undertaken in the WHR2000. Indicator 2 would be the cost-effective coverage of critical interventions. Information on the costs and effectiveness of interventions can help to determine the set of interventions that would maximize population

health for the resources available to a country. The frontier M in Figure 28 (which repeats Figure 1) represents the potential outcome if the most cost-effective mix of interventions had been used for the available resources, or $(d+e+f)$. The ratio of current coverage with that set of interventions to full coverage must be directly related to observed efficiency as defined earlier $\{e/(e+f)\}$. This ratio provides information on the fraction of the potential health gain that has been achieved using the cost-effective set of interventions and can be used as another indicator of efficiency for the MIM estimation.¹¹

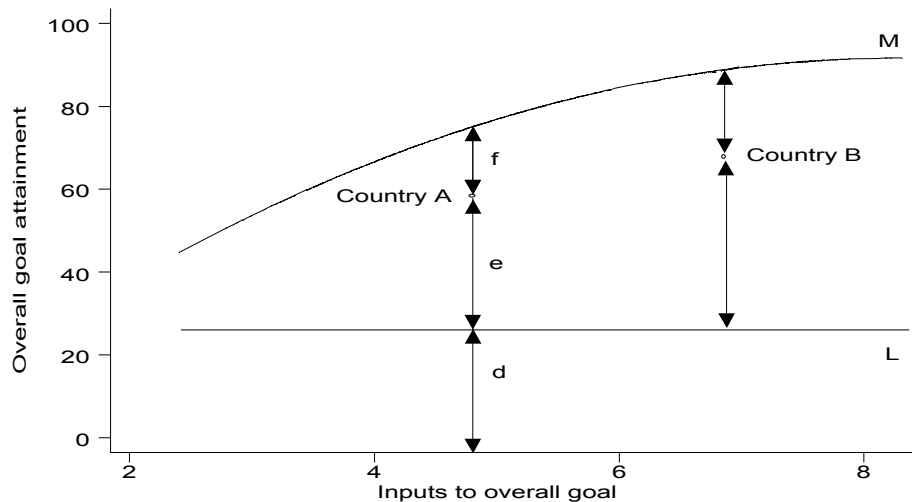


Figure 28. Efficiency in Relation to Cost-Effective Interventions

This model has not yet been estimated because data on effective coverage are not yet available, but we believe that it is possible to do so. It has the potential to make use of all possible sources of information relating to efficiency.

To facilitate this process, WHO is currently developing a data base on the costs and effectiveness of a wide range of interventions through the WHO-CHOICE initiative.¹² Estimates are being made using a standard methodology in 17 regions of the world, to help policy-makers decide which interventions they should be encouraging with the resources available to their systems. Preliminary results for selected interventions relating to blindness are found in Table 14, for four WHO regions. They suggest that some interventions for cataract cost as little as \$5 to gain a disability-adjusted life year. This type of information helps decision makers choose between the vast array of tools and technology available to improve health.

¹¹ Point A in Figure 28 is not strictly comparable to point A in Figure 1 where it was the observed level of goal attainment. That could well be achieved by using a cost-ineffective mix of interventions. However, the ratio of coverage of the cost-effective set to optimal coverage must be directly related to efficiency.

¹² CHOICE stands for **CHO**osing Interventions that are **CO**st-**E**ffective.

Table 14 Cost-effectiveness of blindness control

	AFRO D		AFRO E		AMRO A		AMRO B	
	CER ¹³ (\$)	ICER ¹⁴ (\$)	CER (\$)	ICER (\$)	CER (\$)	ICER (\$)	CER (\$)	ICER (\$)
CATARACT INTERVENTIONS								
ICCE (hospital, low cov)	5	5	5	5	-		14	14
ICCE (hospital, high cov)	6	6	5	6	-		17	18
ECCE-PC-IOL (hospital, low cov)	8		7		128	128	18	
ECCE-PC-IOL (hospital, high cov)	9		9		149	157	21	
ICCE (mobile, high cov)	19		15		-		18	
ECCE-PC-IOL (mobile, high cov)	26		22		-		25	
TRACHOMA INTERVENTIONS								
Mass treatment tetracycline ointment, high coverage (A)	624		867					
Mass treatment tetracycline ointment, low coverage (B)	624		867					
Mass treatment azithromycin, high coverage (C)	583		860					
Mass treatment azithromycin, low coverage (D)	583		860					
Targeted treatment azithromycin (E)	2,802		4,176					
TT surgery (F)	0.23	0.23	0.31	0.31				
Combination of A&F	2		3					
Combination of B&F	1		2					
Combination of C&F	9	583	13	860				
Combination of D&F	1		2					
Combination of E&F	29		43					

¹³ CER = Cost Effectiveness Ratio

¹⁴ ICER = Incremental Cost Effectiveness Ratio

D. Proposals - Efficiency

WHO proposes:

1. The questions of timing and how best to estimate the minimum are complex and we seek the opinions of the PRG. At the same time, the multiple indicator model approach is promising and should be developed further. This requires the proposed work on coverage, discussed in section IV, to continue with the World Health Survey playing an important role.
2. To address the question of timing, two suggestions are made. Firstly, it is necessary to control estimates of the efficiency of producing current period outcomes for historical actions. The simplest way would be use HALE at some time in the recent past, say five years ago, as a controlling variable. Second, we propose that it is worth pursuing the question of incidence HALE as suggested in section V.A. This has the advantage of being much more clearly determined by actions taken this year, but which will not produce an outcome until some time in the future. It would still be necessary to control for that part of incidence HALE determined by actions taken in the past.

VII. MEASUREMENT CHALLENGES

A. Comparable Estimates Using the Best-Available Evidence

1. WHR2000

The philosophy behind the WHR2000 calculations was that decision-makers could not wait until perfect evidence was available. They needed to make decisions now and this requires the best evidence available now. It is rare that no information is available to guide policy, which is the reason that WHO provided estimates for 191 Member States. As shown in the earlier sections of this paper, sometimes these estimates were based on measurements within the countries, and sometimes based on the relationship between the outcome of interest and a set of covariates.

Considerable attention was paid in the WHR2000 to providing uncertainty intervals alongside each point estimate of critical parameters, the first time to our knowledge that an international agency has done so. Considerable attention was also paid to assessing the stability or otherwise of the conclusions using sensitivity analyses, and the results were remarkably robust to changes in assumptions. In the tables of measures that contribute to the overall index of attainment, the WHR2000 showed clearly which estimates were not based on direct population-based data by putting them in italics. This was an additional step to alert readers to the variations in the uncertainty intervals for estimates and represents a greater degree of transparency than is practiced by most international and national statistical agencies in reporting statistical estimates.

The techniques used also made corrections for major known biases in available measurements to improve cross-population comparability. The concept of internal consistency was used as a tool to improve the validity of epidemiological assessments.

We had intended to use the results of existing health and satisfaction surveys to estimate the prevalence of non-fatal health outcomes and health system responsiveness but found that self-

reported health and satisfaction were not comparable across populations. These results could not be used to compare populations within countries or across countries, nor was it correct to aggregate the individual responses to provide national averages. Accordingly, the prevalence of non-fatal health outcomes was based largely on estimates derived from the global burden of disease study and the key informant interviews described earlier were used to estimate responsiveness.

2. Commentaries on the WHR2000

Some of the major issues raised in the consultations have been discussed at length earlier in this document. They include, for example, the recognition that some of the measures used were not comparable across populations.

However, the most common debate has focused on the difference between estimates where the uncertainty intervals are wide and those with tight uncertainty intervals linked largely to sampling error. This was phrased in some consultations as the difference between ignorance and uncertainty. Some commentators argued that there is a fundamental difference between an estimate based on a review of the all available evidence plus the application of plausible methods, and 'proper' measurements as undertaken in many epidemiological studies. The other viewpoint was that at the national level, the distinction between 'real measurement' and estimates is difficult to maintain. All measurements have uncertainty. Any measurement may also be subject to systematic bias and/or is only a proxy for the real quantity of interest. A good estimate is based on a review of all available evidence that is relevant to the quantity being estimated. If necessary, this evidence is then combined with a model or estimation method that captures plausible relationships to develop a most likely estimate and an appropriate uncertainty interval.

This issue is often presented as a conflict between world views. One view, espoused by some epidemiologists, is that assessments should only be reported if they have narrow uncertainty intervals primarily based on sampling error. The alternative is that decision-makers need information on important variables that reflects the available evidence and correctly capture all sources of uncertainty.

3. New Analysis

a) Uncertainty

In an accompanying paper(89) we show that there are many precedents in the UN system for using the Bayesian approach - i.e. estimating key quantities of interest to decision makers for all Member States, based on the available evidence and with varying degrees of uncertainty. These include population numbers and projections, demographic indicators including infant mortality, life expectancies, fertility, child survivorship, migration, marital status and use of family planning, and economic statistics such as GDP. WHO and UNAIDS have also been publishing regular estimates of health-related indicators such as HIV/AIDS mortality and immunization coverage across the world for some years based on similar techniques to those used in the WHR2000.

This conforms to our general approach to describing and estimating uncertainty in quantities of interest. This is to express them as probability distributions using a Bayesian interpretation of probability as expressing uncertainty of an observed or hypothetical event given a set of assumptions about the world. Probability distributions can be used to express uncertainty about empirical or chance quantities representing information about states of the world. These might include epidemiological quantities such as the prevalence of depression in a

particular population, or population preferences such as health state valuations or the weight to be given to different domains of health system responsiveness.

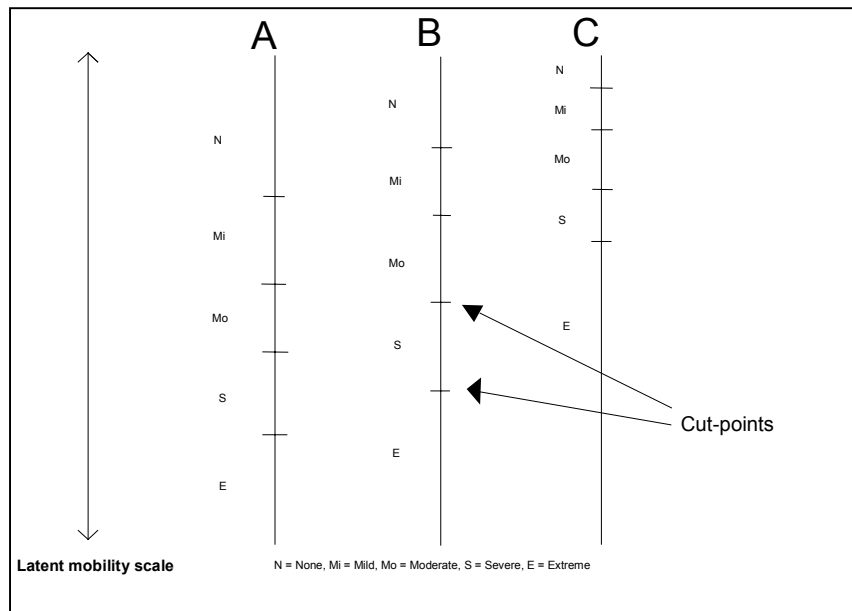
We agree that it would be desirable to have primary data from as many countries as possible and have intensified collaboration with countries to (a) carry out representative population surveys of health status and of health system responsiveness, (b) identify as many recent income and expenditure surveys as possible, critical to the calculations of the fairness of financial contributions, and (c) improve estimates of national accounts and national health expenditures. Details of these developments were described in earlier sections. These new sources of data will reduce the need to use regression techniques to estimate attainment on variables in countries for which ‘direct’ data are unavailable, and the World Health Survey can play a valuable role in this respect.

Major efforts have also been made to develop instruments and methods to improve the cross-population comparability of survey results. As shown earlier, these include the introduction of vignettes into surveys, conducting measured tests in conjunction with surveys, and use of the hierarchical ordered probit model for analysing results.

b) Cross-population comparability

WHO conducted a Multi-country Survey Study in 2000 and 2001, partly to develop and test novel techniques to ensure the cross-population comparability of survey responses.(54;55;90) Self-report categorical responses are often not comparable between different groups of people in cultural or demographic characteristic (i.e., the data exhibit what the psychometricians call *differential item functioning*). To estimate the prevalence of non-fatal health outcomes, for example, people are commonly asked to report their current health, often by domain, according to a limited number of categories of possible response. Figure 29 illustrates the main challenge of using self-reported levels on a domain even when reliability and within-population validity have been well established. For each domain, there is some true or latent scale for a domain. Imagine that Figure 29 shows the latent scale for mobility. In a survey, consider a general question on mobility such as do you have any difficulties walking up stairs. The response categories are “no difficulty”, “mild difficulty”, “moderate difficulty”, “severe difficulty”, and “extreme/cannot do”. The second column in Figure 29 shows for population A, the response category cut-points.

Figure 29. Response category cut-point shifts



We hypothesized that cut-points vary between populations because of different cultural or other expectations for domains of health or responsiveness. Cut-points are also likely to vary within a cultural or socio-demographic group. The cut-points for older individuals may shift as their expectations for a domain diminish with age. Men may be more likely to deny declines in health so that their cut-points may be systematically shifted as compared to women. Contact with health services may influence expectations for a domain and thus shift cut-points.

Comparable measurement across populations requires using the same questions or items in surveys. It also requires explicit strategies to measure the response category cut-points of each item in different populations and socio-demographic groups. One strategy for establishing cross-population comparability is to fix the level of health or responsiveness on a domain and assess variation in the response categories across individuals, groups, and populations. We developed a series of vignettes for this purpose for both health and responsiveness. They described a concrete level of ability or responsiveness on a given domain that individuals were asked to evaluate and this was compared to their self-report on that domain.

For example, self-reported mobility was based on the question: "Overall in the last 30 days how much difficulty did you have with moving around?" Possible responses were:

1=None, 2=Mild, 3=Moderate, 4=Severe, 5=Extreme/Cannot do.

To assess the response category cut-points, each respondent was also asked to assess ability to move around for hypothetical people described with a series of vignettes covering the complete range of mobility. For example, at one extreme was:

Vignette 1: [person's name] is an active athlete who runs long distance races of 20 kilometers twice a week and engages in soccer with no problems.

People were asked to rate the mobility of the hypothetical person according to the same categories used for the self-report. The vignette fixes a given level of mobility so that variation in the response categories is attributable to variation in the response category cut-points alone. Individuals were asked to evaluate a series of vignettes of varying severity across each domain of health and responsiveness. The cut-points they imply can be evaluated using the hierarchical ordered probit (HOPIT) model, discussed later.

A comparable scale for a domain can also be established through calibration tests which capture a domain and which can be implemented in different settings without systematic bias. Such calibration tests are not feasible for a number of domains of health such as pain or affect, nor have we yet developed any for responsiveness. But for some domains of health such as vision, hearing, cognition, and mobility, calibration tests are feasible. In the WHO Multi-Country Study, we used the Snellen's chart to measure visual acuity for distant vision. Since literacy is an issue in cross-national surveys, we chose to use the tumbling E's version of the Snellen's chart for measuring distance vision, and for mobility we used the posturo-locomotion-manual (PLM) test.

For the same level of visual acuity, as measured using the Snellen's eye exam, some people classify themselves as having moderate difficulty, some as severe, and some as having extreme difficulty. The self-reports can be adjusted for these apparent cut-point shifts using the HOPIT model.

The hierarchical ordered probit (HOPIT) model uses information from vignettes and/or measured tests to calibrate self-report responses so as to make them cross-population comparable.⁽⁵⁵⁾ The model is a variant of the standard ordered probit model in that cut-points are allowed to be functions of explanatory variables. Estimation is done by maximum

likelihood methods. There are two components to the likelihood function. In the first component, responses to vignettes are analyzed in order to derive cut-point differences across socio-demographic groups. In a second component of the likelihood function, these estimated cut-points are used to calibrate self-report responses in an interval regression type framework in order to estimate the underlying latent variable.

Figure 30 shows the way that adjustments using HOPIT alter the mean response categories – similar to the example of mobility given in section V. This time, the item is self-care, one of the domains of health. The adjustment shows that mean responses decline much more rapidly with age than unadjusted self reports.

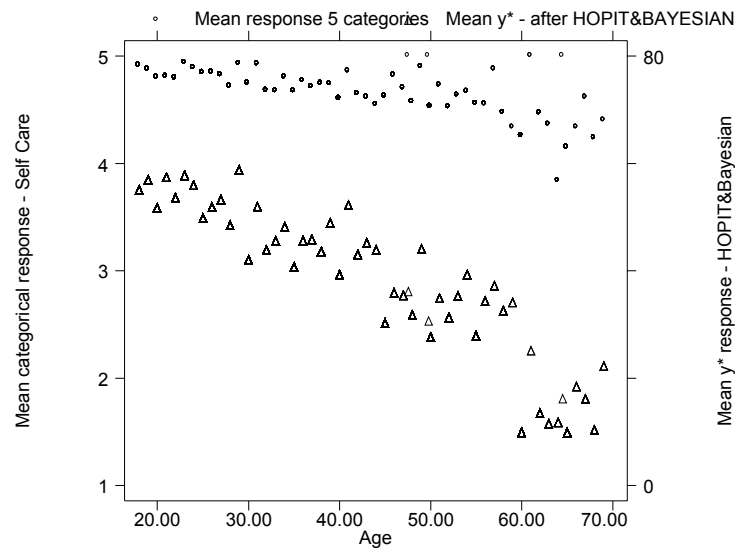


Figure 30. Adjusted and Unadjusted Responses, Self-Care, India

4. Proposals

Policy-makers charged with improving the performance of their health systems must currently do so in the face of a remarkable lack of evidence about the performance of their systems and the impact of the possible ways of organizing their systems on the outcomes of concern to their populations. One of the goals of WHO is to provide a means for them to measure the performance of their systems, to monitor it over time, and to compare their performance with that of other countries. This requires providing decision-makers with the best available evidence currently available, with an assessment of the uncertainty around these estimates.

The methods that are currently available to do this are not perfect, and they will improve with vigorous debate, with peer review and with use. They will also benefit from increased efforts to collect primary data. It is important that the estimates reported by WHO are based on

transparent methods which are replicable. Policy-makers then have the choice to use them in the way they find most appropriate.

WHO proposes to:

- 1. Work with Member States to increase the availability of primary data on key parameters;**
- 2. Continue to estimate and report key quantities of interest, with their uncertainty intervals, for all Member States in the health system performance analysis using transparent and replicable methods.**
- 3. Continue to develop and test methods to ensure the cross-population comparability of survey results.**

B. Data Collection Strategies and The World Health Survey

1. WHR2000

For the most part, data collection for the WHR2000 was *ad hoc*, much of it described in earlier sections of this document. For example, available household expenditure surveys were used to calculate fairness of financial contributions. Demographic and health surveys were the source of most of the information on child mortality and a very intensive effort was undertaken collect the best available secondary data on national health expenditures. Existing data-bases within WHO provided some information on mortality and the prevalence of non-fatal health outcomes.

Only limited primary data collection was undertaken. This involved the web-based survey to obtain the weights for the construction of the overall attainment index, and a country-based key informant survey undertaken for responsiveness.

2. Commentaries on the WHR2000

One of the most common concerns about the WHR was the perceived lack of primary data collection, and the number of countries for which estimates had been made based on covariates. These comments were particularly strong in relation to fairness of financial contributions, responsiveness and health inequalities but they were also raised in relation to estimates of the prevalence of non-fatal health outcomes and even death rates and life tables.

Other comments included:

- key informants are not an appropriate data collection method when data that are representative of population experiences are required;
- the data collected by WHO and required for performance assessment is too onerous and costly to collect, and overlaps with that provided or sought by other agencies.

3. New Developments

a) Data availability

Many of the questions raised in the consultations have been addressed in earlier sections, including the extent to which additional sources of micro data have become available. These questions will not be discussed further in this section, which will focus on a means of obtaining comparable data at relatively low cost, in a way that supplements the activities of the routine health information system.

b) The WHO Multi-Country Survey Study 2000-2001

WHO began its Multi-Country Survey Study late in 2000 as a means of piloting tools to be included in the projected **World Health Survey**.⁽⁹¹⁾ This study developed a common survey instrument with a modular structure for use in nationally representative (adult) general population samples. The main purpose was to develop the methods and instruments necessary to measure key indicators in a way that is valid, reliable and comparable. Method development included comparison of survey modes, questionnaire design, and issues dealing with cross-population comparability. Although the main focus was on the health of individuals in various domains, additional modules were developed on health system responsiveness, health state valuations, overall goal valuations, household health expenditures, mental health, chronic conditions and basic demography.

The health module was based on selected domains of the International Classification of Functioning, Disability and Health (ICF) and was developed after a rigorous scientific review of existing assessment instruments. The survey instrument was developed in multiple languages using cognitive interviews and cultural applicability tests, stringent psychometric tests for reliability (i.e. test-retest reliability to demonstrate the stability of application) and most importantly utilizing novel psychometric techniques for cross-population comparability.

The study was carried out in 61 countries completing 71 surveys (Table 15). Different modes were intentionally used for comparison purposes in each of ten countries. The modes included face-to-face personal interviews in households, each lasting 90 minutes; brief face-to-face interviews (BFTE); computerized telephone interviews (CATI); and postal surveys.

All samples were selected from nationally representative sampling frames with a known probability so as to make estimates based on general population parameters. Sample sizes are also shown in Table 15.

Table 15. Countries participating in the WHO Multi-country Survey Study 2000-2001

	Countries by WHO Region	Survey Mode	Sample Size	No. of countries covered in region	Total sample size
	AFRO				
1	Nigeria	Full length household	5047		
				1	5047
	AMRO				
2	Argentina	BFTF	1555		
3	Canada	Postal	816		
4	Canada	CATI	778		
5	Chile	Postal	2078		
6	Colombia	Full length Household	6019		
7	Costa Rica	BFTF	1508		
8	Mexico	Full length Household	4812		
9	Trinidad and Tobago	Postal	2582		
10	United States of America	Postal	1736		
11	Venezuela	BFTF	1495		
				9	23379
	SEARO				
12	India	Full length Household	5196		
13	Indonesia	Full length Household	9952		
14	Indonesia	Postal	2996		
15	Thailand	Postal	2288		
				3	20432
	EURO				
16	Austria	Postal	2089		
17	Belgium	BFTF	1100		
18	Bulgaria	BFTF	1010		
19	Croatia	BFTF	3000		
20	Czech Republic	Postal	2020		
21	Czech Republic	BFTF	1090		
22	Denmark	Postal	2684		
23	Estonia	Postal	1000		
24	Finland	BFTF	1021		
25	Finland	Postal	2692		
26	France	BFTF	1003		
27	France	Postal	1525		
28	Georgia	Full length Household	9847		
29	Germany	Postal	1123		
30	Greece	Postal	1750		
31	Hungary	Postal	2996		
32	Iceland	BFTF	489		
33	Ireland	BFTF	711		
34	Italy	BFTF	1002		
35	Kyrgyzstan	Postal	2209		
36	Latvia	BFTF	1512		

37	Lithuania	Postal	3498		
38	Luxembourg	CATI	719		
39	Malta	BFTF	500		
40	Netherlands	BFTF	1085		
41	Netherlands	Postal	1211		
42	Poland	Postal	1707		
43	Portugal	BFTF	1001		
44	Romania	BFTF	1053		
45	Russian Federation	BFTF	1601		
46	Slovakia	Full length Household	1183		
47	Spain	BFTF	1000		
48	Sweden	BFTF	1000		
49	Switzerland	Postal	962		
50	Turkey	Full length Household	5194		
51	Turkey	Postal	4524		
52	Ukraine	Postal	1562		
53	United Kingdom of Great	Postal	1996		
				33	71669
	EMRO				
54	Bahrain	BFTF	1609		
55	Cyprus	Postal	1374		
56	Egypt	Full length Household	4486		
57	Egypt	Postal	2771		
58	Iran (Islamic Republic of)*	Full length Household	10,000		
59	Jordan	BFTF	1604		
60	Lebanon*	Full length Household	5000		
61	Lebanon	Postal	2224		
62	Morocco	BFTF	1506		
63	Oman	BFTF	1719		
64	Syrian Arab Republic*	Full length Household	10,000		
65	United Arab Emirates	BFTF	1686		
				10	43979
	WPRO				
66	Australia	Postal	2773		
67	China	Full length Household	9442		
68	China	Postal	2480		
69	New Zealand	Postal	3401		
70	Republic of Korea	Postal	705		
71	Singapore*	Full length Household	5000		
				5	23801
				61	188307

* Surveys currently in progress

The Multi-Country Survey Study enabled us to develop and test ways of ensuring cross-population comparability, to test different survey modes and questions. Many lessons were

learned. Some vignettes did not perform well, for example, with people categorizing them as being indistinguishable from other vignettes. Some questions did not seem to be well understood, and will have to be modified for the World Health Survey. Moreover, postal surveys and telephone interviews had higher non-response rates. In all survey modes, there were problems detected with the within household sampling. On the other hand, the surveys were of very low cost compared to many of the other survey programmes suggesting that occasional surveys can be a cost-effective way of supplementing the data collected by routine health information systems.

c) The World Health Survey

Routine health information systems provide valuable information on many indicators of interest to policy-makers, including the major medical reasons for inpatient stays and outpatient visits, and the sex and age profile of users of the system. However, in many countries they do not provide critical pieces of information that would be valuable to policy such as coverage rates of critical interventions, the effectiveness of service delivery, the responsiveness of health systems, and private sector service costs and utilization. In addition, countries that are likely to be recipients of large increases in funds from international efforts to scale-up health interventions in poor populations require the baseline information necessary to monitor the impact of scaling-up over time.

To fill these gaps in a low cost way which complements the routine HIS, the Director-General has announced the launch of the **World Health Survey** which builds on the experience gained in the WHO Multi-country Survey Study.⁽⁹¹⁾ It collects baseline information on the population and its health. It has modules on different aspects of the health system, including the effective coverage of critical interventions. **It is an independent exercise to health system performance assessment.** However, it offers an ideal platform to gather information on some aspects of goal attainment, inputs and functions using methods and approaches recommended by the Peer Review Group.

Objectives: The World Health Survey seeks to:

- Provide valid, reliable and comparable information at low cost on relevant outcomes, functions and inputs to health systems to supplement the information provided by routine health information systems;
- Enable countries involved in scaling-up of health interventions as a result of increased international concern with poverty and inequalities to assess baseline data and then to monitor if the increase in resources achieves the desired outcomes.

This will also contribute to:

- Building the scientific basis for the development of evidence-based policy advice and to enable policy-makers to monitor if their health systems are achieving the desired goals;
- Providing policy-makers with the evidence they need to adjust their policies, strategies and programmes as necessary.

A Modular Approach: The World Health Survey uses a survey instrument with a common core and a flexible shell of modules covering various components. Policy-makers can choose from the flexible modules in any combination according to their policy needs and add other modules if they wish. Presently, the modules cover key aspects of outcomes to which systems should contribute, inputs to that system, and one aspect of the way systems are functioning – coverage of key interventions:

- *the health of populations*: measuring health in multiple domains;
- *risk factors and their association with health states*: measuring various risk factors such as tobacco, alcohol, physical activity levels, exposure to pollution;
- *the responsiveness of health systems*: whether health systems serve to meet the legitimate expectations of people;
- *coverage, access and utilization of key health services*: such as immunization, treatment of childhood illness, safe motherhood interventions, DOTS for tuberculosis, STD and HIV/AIDS prevention, mental health interventions, etc.;
- *health care expenditures*: how much households contribute to the health system.;
- *household permanent income*: asset and indicator variables which can be used to make a low-cost assessment of household permanent income.

Surveys need not be undertaken each year. The frequency depends on the needs and capacities of the countries.

Technical Support: To be useful for policy, responses should be representative of the population group under consideration. WHO provides advice on sample selection to avoid bias, and ways of introducing quality assurance procedures during survey implementation to ensure that accurate and reliable data are obtained. Depending on the information needs, the mode and the detail needed, recommended sample sizes vary between 1,000 and 10,000 for each country survey. The first phase of the work covers adult populations (i.e. older than 18 year old). A second phase will focus on the health of younger people.

Comparability is the key to policy relevance. Building on the earlier WHO Multi-country Survey Study, modified vignettes have been built in to the modules. In some special cases, performance tests may also be used. Pilots of modified questions and vignettes are currently being undertaken in a number of participating countries.

Interaction with Routine Health Information Systems: The exact form the survey programme takes in a particular country will be developed in individual countries through consultation with policy-makers, particularly those involved in planning the scaling-up of health activities in response to the prospective increase in available resources. It has also been extensively discussed with the people involved in routine health information systems. It is designed to be complementary to their efforts, to ensure periodic data input in a cost-effective way so that important gaps in health information are covered.

Progress: Currently, approximately 70 countries have expressed the desire to participate in the first round of the WHS including a number that are likely to benefit from additional funds becoming available through scaling-up activities. Sustaining the World Health Survey over time presents a number of challenges. For example, it is important to build systems with a long-term goal of integrating surveys into routine health information systems. This would include capacity building within selected countries so that they can sustain a continued survey programme as and when needed. It is also important to carry out scientific studies to improve survey techniques, computerized data collection utilizing geo-coding and small area analysis; and improved data analysis and reporting for international comparisons. Swift dissemination of results will facilitate the service required by the consumers of this information.

d) Other Information Generation Activities

Other key indicators of service provision and resource generation require data that can only be obtained from samples of providers or facilities. These have been

detailed in section IV. For the function of stewardship, instruments are being designed to gather data from samples of stakeholders on the different stewardship domains.

4. *Proposals - Data Collection Strategies*

WHO proposes that:

- 1. The ongoing World Health Survey offers an ideal platform to seek information on the prevalence of health states, health state valuations, responsiveness levels and distributions, household health expenditures, coverage, basic demography and permanent income which can be useful for health systems performance assessment.**
- 2. Special efforts to build consensus on collaborative approaches with other agencies sponsoring or conducting surveys should be made e.g. Demographic and Health Surveys, Living Standards Measurement Surveys, EURO Barometer, and national surveys).**

VIII. INCREASING THE POLICY RELEVANCE OF HEALTH SYSTEM PERFORMANCE ASSESSMENT

A. Communicating Results to Policy-Makers

1. *WHR2000*

The results of the performance assessment exercise in the WHR2000 were presented as:

- attainment on the 5 components of intrinsic goals individually;
- attainment of the composite index;
- efficiency in terms of health and in terms of the composite index.

All scores were presented in rank order, with uncertainty intervals around scores and ranks.

In some cases the uncertainty intervals were wide and in other cases, much narrower. For example, on overall efficiency Malta ranked between 3rd and 7th in 80% of cases, whereas the 80% uncertainty interval around the efficiency rank for Yemen was much wider – from 92nd to 140th.(8)

2. *Commentaries on the WHR2000*

There was mixed feeling at the consultations about the value of publishing overall attainment scores and accompanying rankings. The former topic was discussed in section V. In terms of ranks, some participants argued that the overall attainment and efficiency estimates should not be reported country-by-country. The alternative was to group countries by a number of categories – perhaps high, medium and low. However, other participants saw the value of ranks in focusing the attention of key policy-makers on the health system and its performance.

Other comments were of a procedural nature. For example, people in government argued that countries need to be given the opportunity to comment on the estimates before they are published, that they should be given substantial advance warning before data is released to the media, and decision-makers and the media need to be given more information and assistance on how to respond to performance assessment information. The latter concern applied particularly to the uncertainty intervals around the ranks which were ignored or misunderstood by the media.

3. *New Analysis*

Since the publication of the WHR2000 we have reviewed the experience of other ranking exercises. The Human Development Index (HDI), published yearly by the UNDP in its Human Development Report, is a case in point, as are rankings of hospital performance in the UK and the USA. On one hand, these ranking exercises have raised questions of fairness. Rankings on the HDI or hospital performance do not take into account all possible determinants of outcomes, so can seem unfair to countries or hospitals facing unusual constraints on their performance. For countries this might include colonial history or a recent natural catastrophe. For hospitals, it might be include being located in locations where it is difficult to attract high quality staff.

On the other hand, even though the HDI raised considerable debate initially, it has focused the attention of policy-makers, the media and even the general public on the fact that development is a multi-dimensional concept involving more than income. The index includes educational and health attainment (life expectancy at birth) in addition to income and it is now generally recognized that human development requires more than economic development. Similarly, hospital ranking exercises have focused the attention of policy-makers and the public on the outcomes which people value.

A tentative conclusion is that rankings are not of particular interest to the technical experts required to take the steps necessary to improve performance - although comparisons of country performance with that in a reference group of countries is useful for this purpose. However, rankings provide the means of gaining the attention of the key decision-makers who are in the position to provide more resources for health and to take the necessary actions required to demonstrate a political will to improving the performance of health systems.

4. *Proposals*

WHO proposes that:

The key components of performance assessment should be published as ranked tables with uncertainty intervals around the scores and the ranks.(92) This should be preceded by a consultative process with countries on data sources and the methods used to produce the estimates. Sufficient advance warning would be provided to Member States before publication of the final estimates to allow them to prepare internally for publication, and for them to request technical input from WHO on particular issues or questions of concern to them.

B. Using Health System Performance Assessment to Improve Performance

1. *Background*

Since the publication of the WHR2000, the regional consultations have argued that the links between the measurement of performance and the development of policy requires strengthening. In addition, a substantial number of countries expressed interest in active collaboration with WHO to assess the performance of their own systems and to use the evidence to formulate policies to improve performance. To meet the country requests, the Director-General decided to group these efforts under the rubric of the *Enhancing Health Systems Performance Initiative (EHSPI)* which we also propose to use to meet the suggestions of the regional consultations.(93)

Table 16. Countries Requesting Collaboration with EHSPI

AFRO Chad; Côte d'Ivoire; Malawi; Kenya; Mozambique; Senegal; South Africa; Zimbabwe	EMRO Iran; Morocco; Oman; Jordan; Syria; Tunisia
EURO Czech Republic; Norway Kyrgyzstan; Spain; UK	PAHO Mexico
SEARO Indonesia; Myanmar; Sri Lanka; Thailand	WPRO China, Malaysia; Papua New Guinea

Almost 30 countries have expressed an interest in being actively involved (Table 16) and we are currently working with twenty countries from all income ranges and all WHO regions. They have expressed a mix of reasons for engaging, including

- assessment of the performance of their own health systems, or sub-systems, using the WHO framework;
- assessment of their own performance using better data than was available to WHO;
- development of national skills in the required methodologies;
- seeking support from WHO for the development of health policies and systems using the available evidence;
- contributing to the development of more practical tools for translating evidence into practical policy, particularly related to the four functions;
- the search for greater contact with and opportunities for learning about health systems in other countries.

2. *EHSPI Purpose and Scope*

EHSPI has both national and global objectives

At a national level, the aims are to:

- Enable policy makers to have a better understanding of their health system's performance, and to feed this information into a national policy debate. This requires an assessment of performance and the analysis of the policy implications;
- Link evidence to actions to improve performance;
- Develop greater national capacity to monitor and improve performance.

The country level work interacts with the two global objectives:

- The further development of the conceptual framework and methods;
- Development of a better international evidence-base for policy advice.

Applying the framework at the country level involves measuring attainment in terms of the three intrinsic health system goals, assessing how well the four core health system functions are being carried out, and analysing the policy implications. These steps can be considered to be part of a diagnostic phase. The 'implementation' type activities envisaged by the Initiative refer to specific actions aimed at improving performance would be carried out and tested, with WHO's assistance where requested. In addition, there are also 'reflection' type activities in which people involved in different aspects of performance assessment at the country level can meet together with counterparts in other countries and with WHO. This allows feedback and further development of the framework, indicators and methods, as well as to stronger links to policy.

3. *Diagnosis: Describing and understanding health system performance*

a) Outcomes

Initial work on EHSPI has shown that working closely with countries to carry out their own baseline assessment of attainment on the intrinsic goals is extremely useful for identifying new data sources, for undertaking new data collection where data do not currently exist, and for refining the definition and measurement of indicators. For example, some countries have been particularly interested in testing what mode of survey is the most cost-effective way of obtaining the desired information, so more than one modality has been tested in some countries. Others have provided feedback which has helped to modify the 5 indicators of the intrinsic goals as shown in section V or questions that will be included in the World Health Survey (section VII).

b) Inputs

To date, most attention has focused on improving estimates of national health expenditures in countries lacking national health accounts. There are several regional initiatives supporting National Health Account (NHA) construction, and EHSPI has so far identified and facilitated their support. A *'Producers Guide to National Health Accounts for Low and Middle Income Countries'*, jointly authored by WHO, the World Bank and USAID, will soon be published in English, French and Spanish. We propose to the PRG that this interaction with NHA networks is the most appropriate way to build evidence in this respect.

c) Functions

A number of countries have requested help to measure the performance of the four basic functions in their settings. To this point, the major emphasis has been on defining an indicator of service provision that is more useful for policy than geographical access and a number of countries have agreed to test the new WHO tool to measure effective coverage, defined in section IV.

4. *Implementation: linking evidence to policy*

A number of participant countries have chosen to hold national seminars to introduce a wider range of decision-makers, analysts and researchers to the performance assessment approach, and to discuss preliminary findings from the baseline analysis of inputs, outcomes and functions. These seminars are an important opportunity to take stock and synthesize information from a disparate set of actors, and have in some cases led to the identification of a second set of analytic activities.

In addition, WHO is providing direct policy support to a small number of countries, incorporating the information being generated from these efforts.

5. Sub-national Performance Assessment

Some countries, particularly those with some form of decentralisation, have suggested that the performance assessment framework, used by WHO at the national level, could be helpful in assessing and improving the performance of sub-national units. It could then become a tool for more effective stewardship and management at the national level.⁽⁹⁴⁾ We are currently discussing ways to develop and test a sub-national assessment tool with six countries.

6. Developing Information Systems to Monitor Performance

Several countries have stated from the start that they want to link the baseline assessment with longer term efforts to improve their information systems. This is particularly important if the PRG recommends that the additional indicators of the performance of the different functions of the system (section IV) are routinely measured. It would pose additional strains on health information systems, and makes it all the more important to examine if some of the required information could be obtained by occasional surveys as a supplement to the other activities of a health information system.

7. Building capacity in health system performance assessment and analysis

For health systems performance assessment to be sustainable at the country level, capacity for undertaking both the diagnostic and implementation phases must be built. A variety of strategies can be suggested, ranging from straightforward briefings on using the methods, through technical support to analysts in-country or at WHO, to formal training workshops. All methods have been piloted – for example, this year there was a training workshop in South Africa (in English) for participants from around the world and one for French speakers is scheduled for early 2002 in Côte d'Ivoire.

Proposals

WHO proposes that the EHSPI initiative provides an excellent platform to ensure the policy relevance of HSPA, and for developing national capacities to monitor and improve performance. This will also have the external benefit of contributing to the further development of the tools and methods as well as to contributing to the evidence-base for health policy advice.

Annex Table 1. Regional reporting categories for Global Burden of Disease 2000 project:: WHO regions and 14 subregions.

WHO region	Mortality stratum	WHO Member States
AFRO	D	Algeria, Angola, Benin, Burkina Faso, Cameroon, Cape Verde, Chad, Comoros, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Madagascar, Mali, Mauritania, Mauritius, Niger, Nigeria, Sao Tome And Principe, Senegal, Seychelles, Sierra Leone, Togo
AFRO	E	Botswana, Burundi, Central African Republic, Congo, Côte d'Ivoire, Democratic Republic Of The Congo, Eritrea, Ethiopia, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, South Africa, Swaziland, Uganda, United Republic of Tanzania, Zambia, Zimbabwe
AMRO	A	Canada, United States Of America, Cuba
AMRO	B	Antigua And Barbuda, Argentina, Bahamas, Barbados, Belize, Brazil, Chile, Colombia, Costa Rica, Dominica, Dominican Republic, El Salvador, Grenada, Guyana, Honduras, Jamaica, Mexico, Panama, Paraguay, Saint Kitts And Nevis, Saint Lucia, Saint Vincent And The Grenadines, Suriname, Trinidad And Tobago, Uruguay, Venezuela
AMRO	D	Bolivia, Ecuador, Guatemala, Haiti, Nicaragua, Peru
EMRO	B	Bahrain, Cyprus, Iran (Islamic Republic Of), Jordan, Kuwait, Lebanon, Libyan Arab Jamahiriya, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Tunisia, United Arab Emirates
EMRO	D	Afghanistan, Djibouti, Egypt, Iraq, Morocco, Pakistan, Somalia, Sudan, Yemen
EURO	A	Andorra, Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Slovenia, Spain, Sweden, Switzerland, United Kingdom
EURO	B	Albania, Armenia, Azerbaijan, Bosnia And Herzegovina, Bulgaria, Georgia, Kyrgyzstan, Poland, Romania, Slovakia, Tajikistan, The Former Yugoslav Republic Of Macedonia, Turkey, Turkmenistan, Uzbekistan, Yugoslavia
EURO	C	Belarus, Estonia, Hungary, Kazakhstan, Latvia, Lithuania, Republic of Moldova, Russian Federation, Ukraine
SEARO	B	Indonesia, Sri Lanka, Thailand
SEARO	D	Bangladesh, Bhutan, Democratic People's Republic Of Korea, India, Maldives, Myanmar, Nepal
WPRO	A	Australia, Japan, Brunei Darussalam, New Zealand, Singapore
WPRO	B	Cambodia, China, Lao People's Democratic Republic, Malaysia, Mongolia, Philippines, Republic Of Korea, Viet Nam Cook Islands, Fiji, Kiribati, Marshall Islands, Micronesia (Federated States Of), Nauru, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu

IX. REFERENCE LIST

Many of the papers that appear in the reference list were prepared for the Peer Review. For full details of the broader literature consulted in the preparation, please refer to the background papers in question.

- (1) Maynard A, Bloor K. Health care reform: informing difficult choices. *International Journal of Health Planning and Management* 1995; 10(4):247-264.
- (2) Collins C, Green A, Hunter D. Health sector reform and interpretation of Policy context. *Health Policy* 1999; 47(1):69-83.
- (3) Hussein AK, Mujinja PG. Impact of user charges on government health facilities in Tanzania. *East African Medical Journal* 1997; 74(12):751-757.
- (4) Feldman R. The ability of managed care to control health care costs: how much is enough? *Journal of Health Care Finance* 2000; 26(3):15-25.
- (5) Moore M. Public sector reform; downsizing, restructuring, improving performance 1996. Geneva, Switzerland, World Health Organization. Forum on Health Sector Reform Discussion Paper 7.
- (6) Durham G, Kill B. Public health funding mechanisms in New Zealand. *Australian Health Review* 1999; 22(4):100-112.
- (7) Ron A. NGOs in community health insurance schemes: examples from Guatemala and the Philippines. *Social Science and Medicine* 1999; 48(7):939-950.
- (8) WHO. The World Health Report 2000. Health systems: improving performance. 2000. Geneva, WHO.
- (9) WHO, Regional Office for the Americas. Report on Regional Consultation on Health Systems Performance Assessment (Washington, DC, June 2001). 2001. Washington DC, WHO Regional Office for the Americas.
- (10) WHO, Regional Office for Africa. Report on Regional Consultation on Health Systems Performance Assessment (Harare, July 2001). 2001. Harare, WHO Regional Office for Africa.

- (11) WHO, Regional Office for the Eastern Mediterranean. Report on Regional Consultation on Health Systems Performance Assessment (Lebanon, July 2001). 2001. Cairo, WHO Regional Office for the Eastern Mediterranean.
- (12) WHO, Regional Office for Europe. Report on Regional Consultation on Health Systems Performance Assessment (Copenhagen, September 2001). 2001. Copenhagen, WHO Regional Office for Europe.
- (13) WHO, Regional Office for South-East Asia. Report on Regional Consultation on Health Systems Performance Assessment (New Delhi, June 2001). 2001. New Delhi, WHO Regional Office for South-East Asia.
- (14) WHO, Regional Office for the Western Pacific. Report on Regional Consultation on Health Systems Performance Assessment (Manila, July 2001). 2001. Manila, WHO Regional Office for the Western Pacific.
- (15) WHO. Report on Technical Consultation on Measurement of Efficiency of Health Systems (New Orleans, January 2001). 2001. Geneva, WHO.
- (16) WHO. Report on Technical Consultation on Concepts and Methods for Measuring the Responsiveness of Health Systems (Geneva, September 2001). 2001. Geneva, WHO.
- (17) WHO. Report on Technical Consultation on Effective Coverage in Health Systems (Rio de Janeiro, September 2001). 2001. Geneva, WHO.
- (18) WHO. Report on Technical Consultation on Statistical Strategies for Cross-Population Comparability (Boston, October 2001). 2001. Geneva, WHO.
- (19) WHO. Report on Technical Consultation on Fairness of Financial Contribution (Geneva, October 2001). 2001. Geneva, WHO.
- (20) WHO. Report on Technical Consultation on Health Inequalities (Geneva, November 2001). 2001. Geneva, WHO.
- (21) WHO. Report on Technical Consultation on Stewardship (Geneva, September 2001). 2001. Geneva, WHO.
- (22) WHO. Report of WHO Meetings of Experts on Measuring and Summarizing Health (various). 2001. Geneva, WHO.

- (23) Murray CJL, Frenk J. A framework for assessing the performance of health systems. *Bulletin of the World Health Organization* 2001; 78(6):717-732.
- (24) Murray CJL, Frenk J, Evans DB. Reflections on an elaborated framework. 2001.
- (25) Poullier JP, Kawabata K, Hernandez P. National Health Accounts: concepts and results for 1991 countries in 1998. 2001.
- (26) Poullier JP, Hernandez P, Kawabata K, Savedoff W, Indikadahena C, Zeramdini R. Global and national spending patterns: results from National Health Accounts. 2001.
- (27) WHO. World Health Report 2001. Mental health: new understanding, new hope. Geneva: WHO, 2001.
- (28) Savedoff W, Carrin G, Kawabata K, Mechbal A. The financing function: concepts and proposals for monitoring. 2001.
- (29) Adams O, Shengelia B, Larizgoitia, I, Issakov A, Siem F, Jam T. Provision of personal and non-personal health services: proposals for monitoring. 2001.
- (30) Shengelia B, Adams O, Thieren M, Berckmans Y, Kwankam Y, Murray CJL. Measuring the coverage of critical interventions through household surveys. 2001.
- (31) Adams O et al. Data collection strategies to monitor provision, financing and resource generation of the public and private sectors of the health system.
- (32) Murray CJL, Shengelia B, Gupta N, Moussavi S, Thieren M. Comparing routine reporting DTP3 coverage and household survey assessments. 2001.
- (33) Murray CJL, Moussavi S, Gupta N, Adams O, Shengelia B. Inequality in coverage: concepts and measurement strategies. 2001.
- (34) Adams O, Shengelia B, Issakov A, Dal Poz M, Stilwell B. Human, physical and intellectual resource generation – proposals for monitoring. 2001.
- (35) UNICEF. Problems and priorities regarding recurrent costs. Executive Board EaSC, (editor). Policy Review. 1998.

- (36) Davies P, Mechbal A, Travis P, Kawabata K, Egger D. Stewardship: proposals for monitoring critical domains. 2001.

- (37) Institute of Medicine. Summarizing Population Health: Directions for the Development and Application of Population Metrics. Washington, DC: National Academy Press, 1998.

- (38) Murray CJL, Lopez AD. The Global Burden of Disease, vol. 1. Cambridge MA: Harvard University Press, 1996.

- (39) Katz S, Branch LG, Branson MH, Papsidero JA, Beck JC, Greer DS. Active life expectancy. *The New England Journal of Medicine* 1983; 309(20):1218-1224.

- (40) Bone MR. International efforts to measure health expectancy. *Journal of Epidemiology and Community Health* 1992; 46:555-558.

- (41) Bronnum HH. Trends in health expectancy in Denmark, 1987-1994. *Danish Medical Bulletin* 1998; 45:217-221.

- (42) Crimmins EM, Saito Y, Ingengneri D. Trends in disability-free life expectancy in the United States, 1970-90. *Population and Development Review* 1997; 23(3):555-557.

- (43) Mutafova M, van-de-Water HP, Perenboom RJ, Boshuizen HC, Maleshkov C. Health expectancy calculations: a novel approach to studying population health in Bulgaria. *Bulletin of the World Health Organization* 1997; 75(2):147-153.

- (44) Sihoven AP, Kunst AE, Lahelma E, Valkonen T, Mackenbach JP. Socio-economic inequalities in health expectancy in Finland and Norway in the late 1980s. *Social Science and Medicine* 1998; 47:303-315.

- (45) Valkonen T, Sihvonen AP, Lahelma E. Health expectancy by level of education in Finland. *Social Science and Medicine* 1997; 44:801-808.

- (46) Murray CJL, Lopez AD. Regional patterns of disability-free life expectancy and disability-adjusted life expectancy: Global Burden of Disease Study. *Lancet* 1997; 349(9062):1347-1352.

- (47) Mathers C, Sadana R, Salomon JA, Murray CJL, Lopez AD. World Health Report 2000: Healthy Life expectancy in 191 countries, 1999. *Lancet* 2001; 357:1685-1691.

- (48) Mathers C, Sadana R, Salomon JA, Murray CJL, Lopez AD. Estimates of DALE for 191 countries: methods and results. 2000. Geneva, WHO. GPE Discussion Paper No.16.
- (49) Sadana R, Mathers C, Lopez AD, Murray CJL. Comparative analysis of more than 50 household surveys on health status. 2000. Geneva, WHO. GPE Discussion Paper No.15.
- (50) Murray CJL, Ferguson B, Lopez AD, Guillot M, Salomon JA, Ahmad O. The modified logit lifetable system: principles, empirical validation, and application. 2001.
- (51) Lopez AD, Ahmad O, Guillot M, Inoue M, Ferguson B. Life tables for 191 countries for 2000: data, methods, results. 2001.
- (52) Sadana R, Salomon JA, Tandon A, Chatterji S, Murray CJL. Health state vignettes: design, empirical analysis and critical assessment.2001.
- (53) Tandon A, Chatterji S, Ustun B, Salomon JA, Murray CJL. Cross-validation of cutpoint estimation using measured tests and vignettes: the case of vision.2001.
- (54) Salomon JA, Tandon A, Murray CJL. Using vignettes to improve cross-population comparability of health surveys: concepts, design and evaluation techniques. 2001.
- (55) Tandon A, Murray CJL, Salomon JA. Statistical models for enhancing cross-population comparability. 2001.
- (56) Sadana R, Tandon A, Chatterji S, Ustun B, Murray CJL. Describing population health in six domains: comparable results from 65 household surveys. 2001.
- (57) Salomon JA, Murray CJL, Ustun B, Chatterji S. Health state valuations in summary measures of population health. 2001.
- (58) Fries JF. Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 1980; 303(3):130-5.
- (59) Robine JM, Mathers C, Brouard N. Trends and differentials in disability-free life expectancy: concepts, methods and findings. In: Caselli G, Lopez AD (editors). *Health and mortality among elderly populations*. 1996. Clarendon Press, Oxford, pp182-201.

- (60) Gakidou EE, Murray CJL, Frenk J. Defining and measuring health inequality: an approach based on the distribution of health expectancy. *Bulletin of the World Health Organization* 2000; 78(1):42-54.
- (61) Gakidou EE, King G. Determinants of inequality in child survival: results from 40 countries.
- (62) Gakidou EE, Tandon A, King G, Murray CJL. Measurement methods for inequality in the risk of adult mortality. 2001.
- (63) Almeida C, Braverman P, Gold MR, Szwarcwald C, Ribeiro JM, Miglionico A, Millar JS, Porto S, do Rosario Costa N, Rubio VO, Segall M, Starfield B, Travassos C, Uga A, Valente J, Viacava F. Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet* 2001; 357(9268):1692-1697.
- (64) Williams A. Science or Marketing at WHO? A commentary on "World Health 2000". *Health Economics* 2001; 10(2):93-100.
- (65) Wagstaff A. Measuring equity in health care financing: reflections on and alternatives to the WHO's Fairness Financing Index. 2001.
- (66) Wagstaff A, van Doorslaer E. Paying for health care: quantifying fairness, catastrophe and impoverishment, with applications to Vietnam 1993-1998. Paper presented at the International Health Economics Association (Ihea) Conference July 2001.
- (67) Xu K, Hanvoravongchai P, Murray CJL, Kawabata K, Ortiz JP, Zeramdini R, Klavus J. Household health system contributions and capacity to pay: definitional, empirical and technical challenges. 2001.
- (68) Xu K, Murray CJL, Carrin G. Summary measures of the distribution of household financing contributions to health. 2001.
- (69) Murray CJL, Xu K, Hanvoravongchai P. Decomposing the fairness in financial contribution index into the contribution of vertical inequality, severe horizontal inequality and mild to moderate horizontal inequality. 2001.
- (70) Murray CJL, Xu K, Kawabata K, Hanvoravongchai P, Klavus J, Zeramdini R. Assessing the impact of health system contributions on households: changes in income distribution, changes in poverty, the household distribution of sacrifice, and catastrophic payments. 2001.

- (71) Xu K, Mechbal A, Zeramdini R, Hanvoravongchai P, Kawabata K. Catastrophic payments: cross-country and within country determinants. 2001.
- (72) Xu K, Murray CJL, Zeramdini R, Hanvoravongchai P, Kawabata K, Carrin G. Explaining cross-country variation in the fairness in financial contribution. 2001.
- (73) Blendon R, Kim Benson JM. The Public versus the World Health Organization on Health System Performance: who is better qualified to judge health care systems: public health experts or the people who use the healthcare? *Health Affairs* 2001; (20)3:10-20.
- (74) Murray CJL, Valentine N, Kawabata K. People's experience versus people's expectations. Satisfaction measures are profoundly influenced by people's expectations, say these WHO researchers. *Health Affairs* 2001; 20(3):21-4.
- (75) Ustun B, Chatterji S, Villanueva M, Bendib L, Celik C, Sadana R, Valentine N, Mathers C, Ortiz JP, Tandon A, Salomon JA, Cao Y, Wan Jun X, Murray CJL. The WHO Multicountry household survey on health and responsiveness 2000-2001. 2001.
- (76) Valentine N, De Silva A, Salomon JA, Murray CJL, Kawabata K, Ortiz JP. Responsiveness vignettes: design and empirical assessment. 2001.
- (77) Valentine N, Liu B, Ortiz JP, Murray CJL. Classical psychometric assessment of responsiveness instrument in the WHO Multi-country study of health and responsiveness. 2001.
- (78) Valentine N, Ortiz JP, Liu B, Tandon A, Kawabata K, Poe R, Murray CJL. Levels of responsiveness in eight domains for outpatient and inpatient experiences in 60 national country surveys. 2001.
- (79) Valentine N, Salomon JA. Weights for Responsiveness domains: analysis of country variation in 57 national country surveys. 2001.
- (80) Ortiz JP, Valentine N, Gakidou EE. Measuring inequality in responsiveness. 2001.
- (81) Navarro V. Assessment of the World Health Report 2000 *Lancet* 2001; 357(9241):1598-1601.
- (82) Oswald Cruz Foundation - Ministry of Health. Reports of the workshop "Health Systems Performance - The World Health Report 2000". Rio de Janeiro, 14-15 December 2000.

- (83) Gakidou EE, Evans DB, Murray CJL. Empirical assessment of the importance of three health system goals: results from household surveys in 50 countries. 2001.
- (84) Lauer JA, Evans DB. Measuring health system attainment: the impact of variability in the importance of social goals. 2001.
- (85) Evans DB, Murray CJL, Tandon A, Lauer J. The comparative efficiency of national health systems: a global cross-national econometric analysis. *British Medical Journal* 2001; 323(7308):307-310.
- (86) Mckee M. Measuring the efficiency of health systems. The World Health Report sets the agenda, but there's still a long way to go. *British Medical Journal* 2001; 323(7308):295-6.
- (87) Jamison DT, Sandhu ME. Global health: WHO ranking of health system performance. *Science* 2001; 293(5535):1595-1596.
- (88) Lauer JA, Tandon A, Evans DB, Murray CJL. Determinants of performance: second stage efficiency analysis. 2001.
- (89) Murray CJL, Mathers C, Salomon JA, Lopez AD. Evidence for health policy: dealing with missing data, uncertainty, and ignorance. 2001.
- (90) Murray CJL, Tandon A, Salomon JA, Mathers C, Sadana R. Cross population comparability of evidence for health policy. 2001.
- (91) Ustun B, Murray CJL. World Health Survey: objectives, design, and modules. 2001.
- (92) Adams O, Mechbal A, et al. Communicating performance results to health system actors and the public. 2001.
- (93) Travis P, Mechbal A. EHSPi status report. 2001.
- (94) Travis P, Mechbal A, Murray CJL. Sub-national performance assessment: objectives, challenges and strategies. 2001.