

REPORT OF THE
SCIENTIFIC PEER REVIEW GROUP
ON
HEALTH SYSTEMS PERFORMANCE ASSESSMENT



WORLD HEALTH ORGANIZATION

May 2002

This report contains the collective views of an international group of experts, and does not necessarily represent the decisions or the stated policy of the World Health Organization.

This document is not a formal publication of the World Health Organization and all rights are reserved by the Organization. The document may, however, be freely reviewed, abstracted, reproduced or translated, in part or in whole, but not for sale or for use in conjunction with commercial purposes. The views expressed in documents by named authors are solely the responsibility of those authors.

© World Health Organization, 2002

**REPORT OF THE
SCIENTIFIC PEER REVIEW GROUP
ON
HEALTH SYSTEMS PERFORMANCE ASSESSMENT**

Sudhir Anand, India (Chair)
Walid Ammar, Lebanon
Timothy Evans, Canada
Toshihiko Hasegawa, Japan
Katarzyna Kissimova-Skarbek, Poland
Ana Langer, Mexico
Adetokunbo O. Lucas, Nigeria
Lindiwe Makubalo, South Africa
Alireza Marandi, Iran
Gregg Meyer, USA
Andrew Podger, Australia
Peter Smith, United Kingdom
Suwit Wibulpolprasert, Thailand



WORLD HEALTH ORGANIZATION

TABLE OF CONTENTS

Acknowledgements	i
List of Acronyms	ii
Executive Summary	iii
I. The Framework for HSPA	1
II. Health System Inputs	11
III. Resource Generation Function	19
IV. Service Provision Function	27
V. Financing Function	33
VI. Stewardship Function	41
VII. Average Level of Population Health	51
VIII. Health Inequality	61
IX. Responsiveness - Level and Distribution	71
X. Fairness in Financial Distribution	83
XI. Composite Goal Attainment	93
XII. Data Quality and Data Collection Strategies	103
XIII. Cross-Population Comparability	109
XIV. Efficiency	117
XV. Enhancing Policy Relevance	129
Appendix A - SPRG List of Members	139
Appendix B - WHO Meetings and Consultations on Health System Performance Assessment	141

EXECUTIVE SUMMARY

The Scientific Peer Review Group (SPRG) on Health Systems Performance Assessment (HSPA) was set up by the Director-General at the end of October 2001. The list of 13 members of SPRG is attached as Annex 1. Its terms of reference were:

- To review the scientific merit of methods proposed by the WHO Secretariat for the next round of HSPA, building on the suggestions made in the technical, regional and country consultations, in ongoing research and the general academic debate;
- To propose revisions, as necessary, to the methods that improve their scientific merit, and work with the WHO Secretariat to assess the feasibility and impact of any revision;
- To advise the Director-General of the scientific merit of the final methods emerging from this process.

SPRG met for the first time in December 2001, and prepared an interim report that was presented to the WHO Executive Board in January 2002. The Group had two subsequent meetings in February and April 2002. Each of the three SPRG meetings was attended in person by at least nine members, with most of the others participating via a video-conference or tele-conference link.

This is the final report of SPRG, presented to the Director-General in May 2002. The report has been prepared with input from every member, and the conclusions and recommendations in it are unanimous.

SPRG considers that the objectives of the HSPA initiative are valid, and that the provision of comparative data on health-system characteristics is a vital component of securing health-system improvements. In its deliberations SPRG has therefore sought to apply the following overarching criterion to inform its recommendations: that all future HSPA activity should be judged by the extent to which it effects an improvement in health systems performance worldwide, particularly in countries with low levels of attainment.

SPRG welcomes the opportunity it has been given to contribute to the HSPA process. WHR 2000 made an important breakthrough in seeking to provide an integrated quantitative assessment of health systems performance, and bringing the topic of health-system performance to the attention of policy makers worldwide.

SPRG considers that many of the important issues that have been raised in the public debate about HSPA are strategic policy concerns rather than scientific concerns. The strategic concerns may be matters on which WHO will need to determine a policy, but are in general beyond the remit of SPRG. The Group has therefore sought wherever possible to focus only on the scientific aspects of HSPA.

Within the limited time and resources at its disposal, SPRG has sought to review the scientific evidence from five main sources:

1. Published and unpublished documents and presentations by WHO staff.
2. The reports of the WHO regional consultations and technical workshops.
3. The reports of the WHO meetings of experts.
4. Commentaries by national governments and agencies.
5. Published literature in peer-reviewed journals and unpublished working papers by external commentators.

In addition, during the review process, the Group has been open to considering comments and criticisms received in the form of personal communications from various quarters -- researchers, academics, and professionals in the public policy area.

SPRG wishes to congratulate and thank the WHO Evidence and Information for Policy (EIP) Cluster for the breadth and quality of the materials presented. An enormous volume of material has been made available to SPRG, and members of all Departments in the Cluster were unfailingly helpful in making themselves available and responding to requests for clarification and additional material. Without this responsiveness, our job would have been impossible.

The responsiveness of the EIP staff was an immensely encouraging aspect of the SPRG process. Paradoxically, however, it did generate problems for SPRG, in the sense that the Group frequently found itself commenting on what one member referred to as a 'moving target'. WHO proposals were refined over the course of the review process, leading to the production of numerous new working papers as the review process progressed.

The general approach adopted by SPRG has been to follow the template set out by the WHO Secretariat in its Summary Document "Proposed Strategies for Health Systems Performance Assessment" (in *Background Documentation for Scientific Peer Review Group Meeting, Geneva, 7-8 December 2001*). This included 15 topic areas, which correspond to the sections set out in the main body of this report. For each topic we have sought to describe the approach

taken in WHR 2000, summarize the criticisms that WHR 2000 attracted, outline the subsequent response by WHO, and put forward our comments and recommendations.

In reviewing the material made available, SPRG also developed some overarching recommendations that apply across a wide range of HSPA activity. They can be summarized as follows.

1. The development of local capacity to provide and interpret comparative data is essential to the effectiveness and sustainability of HSPA. It is also likely to be a highly cost-effective use of HSPA resources. Attention should be given to mechanisms of developing capacity at regional and country level, through processes such as "Enhancing Health Systems Performance Initiative" (EHSPI), promoting regional networks, nurturing academic networks, implementing training courses, and encouraging active user engagement.
2. HSPA should be a dynamic, interactive process in which users and other stakeholders are actively involved at both conceptual and implementation stage. HSPA may induce beneficial responses within nations, but unless carefully designed it has the risk of being ineffective, or of inducing undesirable outcomes, such as lack of attention to long-term health system goals. Therefore, in order to achieve its goals, it is imperative that HSPA has a positive influence on Ministries of Health and other key stakeholders. WHO should consider whether it is possible systematically to evaluate the impact of HSPA on Member States.
3. WHO should use rigorous scientific methods in developing and implementing new measurement tools. WHR 2000 was criticized for inadequate engagement with, and recognition of the contributions of, experts in the field. SPRG recognizes that, like all scientific endeavours, the methods will evolve over time. The Group considers it is imperative that future methodology is developed in collaboration with relevant outside experts, and welcomes the recent consultative processes initiated by WHO. Mechanisms to secure expert engagement include expert panels, independent peer review, and secondments to and from relevant institutions. SPRG also encourages WHO to work closely with other international bodies with expertise in this area, such as OECD.
4. Numerous technical judgements have to be made at every stage of the HSPA methodology. There is a need for WHO to prepare a careful audit trail of such judgements, and to make this available for public scrutiny.

5. Notwithstanding the need for scientific rigour, the methods used should be as simple as possible, subject to being fit-for-purpose. HSPA introduces many new concepts and methodologies that are challenging for governments and other stakeholders, and any unnecessary complexity is a serious impediment to communication. The final product should be a set of scientifically sound, practical, user-friendly tools that achieve the objectives of HSPA in enhancing health-system performance.
6. The research function implicit in HSPA should be distinguished carefully from operational implementation. Methods and data sources should be robust, credible, sustainable, and cost-effective before full implementation. In the meantime, they should be presented as work-in-progress, and should be developed using the collaborative and open research process advocated above. It may be helpful for WHO to develop explicit criteria against which it can evaluate initiatives being considered for implementation within HSPA.
7. Great care should be taken with the dynamic aspects of health-system performance. Many actions, particularly in the domain of public health, may have effects on outcomes only after a considerable time lag, and the methodologies used should reflect this complication. Furthermore, policy makers are naturally concerned with national trends over time. Therefore, as methodologies and datasets change, there will arise an important need to ensure that consistent time-series of data are made available to countries.
8. There is an urgent need to improve the quality and continuity of the data on which HSPA is based. Detailed recommendations are given in relevant sections of this report. Particularly important means to this end will include nurturing the development of sustainable health-information systems within countries, development of user skills and capacity, implementation of new data collection tools, and use of cost-effective quality assurance instruments.
9. The World Health Survey (WHS) is a particularly important new development within HSPA. SPRG welcomes the introduction of WHS, acknowledging its potential to inform diverse constituencies concerned with the performance of health systems. SPRG recommends that developmental work to ensure its effectiveness and reliability must continue over time, and its detailed recommendations are given in Section XII. The Group noted that WHS should wherever possible build on existing survey platforms, be useful for local purposes, and not put an unsustainable burden on local capacity for data collection. SPRG also notes that WHS is likely to be of

greatest benefit in countries with poor information systems and low levels of health-system attainment. It therefore recommends that WHO gives priority in WHS and its implementation to the needs of such countries.

10. SPRG welcomes the WHO proposal to develop a parsimonious set of indicators related to the financing, service provision and resource generation functions (in the form of a 'dashboard' approach). The Group offers detailed recommendations in the relevant sections of this report, but considers that the development of a set of reliable, valid, cost-effective, and comparable indicators of health-system functions is an urgent requirement to enhance the usefulness of HSPA.
11. WHO should consider publishing an HSPA report card for every country, which offers a diagnostic tool in the form of a commentary on issues such as measured performance and prospects for improvement. The exact content of these should be determined in consultation with Member States, and should reflect the criterion of cost-effective use of WHO resources. SPRG suggests that the report cards could include a commentary on data quality and assumptions, on progress made since the last HSPA, and on aspects of performance that appear to merit further investigation.
12. SPRG has examined carefully the role of "league tables" of health-system performance within the HSPA process. It considers that the decision as to whether or not to publish such league tables is ultimately a policy and strategic decision for WHO rather than a technical issue. However, there were serious technical questions raised about the WHR 2000 methodology relating to the weights used in the composite index, the scaling of the component indicators, and the treatment of missing data. These criticisms have been documented in the subsequent sections of this report, which also give our detailed response to the WHO proposals for addressing these criticisms.

The following sections report the results of our detailed scrutiny of each of the 15 topic areas. They bear testimony to the extraordinary breadth and richness of the agenda unleashed by WHR 2000. Within the limited time and resources available, SPRG has found it extremely challenging to cover all the issues raised. We nevertheless hope that the treatment of the topics can serve as an adequate basis for informing progress on HSPA in the near future. We have sought to reflect the major issues raised in WHR 2000, and have made numerous detailed recommendations. The main messages from our review are now briefly noted under the 15 headings.

- I. SPRG broadly endorsed the **framework for HSPA**, but in Section I makes some detailed comments designed to clarify and refine the concept.
- II. SPRG noted the extensive work that has already gone into the development of measures of **health system inputs**, in the form of the national health accounts. Section II offers a large number of detailed observations and suggestions for improvement.
- III. SPRG welcomes the attention now placed on the **resource generation function**, but considers current WHO thinking to be at an early stage of development. Section III offers some preliminary observations, but we recommend that this topic should be developed in full consultation with relevant users and experts.
- IV. SPRG considers that the **service provision and coverage function** is particularly important for nations seeking to understand the reasons for their measured level of health-system performance. In particular, WHO has started to develop an ambitious methodology that contains promising implications for operational measurement. However, the methodology will need continued elaboration, refinement, and clarification.
- V. We agree that WHO should continue to develop operational measures relating to the **financing function**. There is a need for research that provides evidence on how the financing function affects health-system performance.
- VI. SPRG welcomes the emphasis on the **stewardship function** in WHR 2000. Although it considers that the measurement of stewardship poses serious challenges and could be a sensitive area, SPRG suggests that WHO should develop and test the proposed new tools.

- VII. Methodology for the measurement of **average level of population health** is relatively advanced. A number of technical issues have been raised concerning the estimation of Health-Adjusted Life Expectancy (HALE), and these are treated in detail in Sections VII and XIII.
- VIII. The concept and measurement of **health inequality** have generated some of the most contentious debates arising from WHR 2000. This HSPA goal poses epistemological as well as policy challenges, and introduces serious practical measurement difficulties. SPRG is not aware of any current data sources that allow international measurement of inequality in the chosen measure for 'average level of population health', HALE (rather than inequality in child survival to age 2 as used in WHR 2000). Hence, SPRG recommends that the 'pure health inequality' approach to examining 'health inequalities' should be developed further at both a methodological and statistical level, and acknowledges that measuring 'socio-economic inequalities in health' is a valuable additional approach.
- IX. The treatment of **level and distribution of responsiveness** in WHR 2000 was weak, relying on Key Informant surveys administered in only a fraction of Member States. The introduction of the World Health Survey will for the first time provide population-based information on responsiveness. However, further work is required to define the concept of responsiveness and identify its importance in different cultural settings and at different stages of development.
- X. The concept and measurement of the **fairness of financial contributions** have attracted a great deal of debate since WHR 2000 was published. Although there are some as yet unsettled technical questions, many of the concerns expressed in the debate relate to policy choices that WHO will have to make and defend.
- XI. SPRG considers that the decision on whether or not to continue to publish a **composite index** of health-system performance is ultimately a policy decision for WHO rather than a technical issue. However, there were serious technical questions raised about the WHR 2000 methodology, which are addressed in Section XI.
- XII. Data inadequacies were a chief source of concern in commentaries on WHR 2000. In response, WHO has launched a major initiative on **data quality and data collection strategies**, including the World Health Survey. As noted above, SPRG welcomes this development, but has raised serious concerns that are detailed in Section XII. SPRG

recommends that WHO makes intensive efforts to obtain household survey data in as many countries as possible, and reduces the need to estimate missing data to a minimum.

- XIII. SPRG considers that the methods proposed to achieve **cross-population comparability** are necessary and innovative. The methodology represents a major advance in comparing self-reported survey responses of different population groups (countries). The methods are still at a developmental stage, and require extensive further testing for robustness.
- XIV. SPRG acknowledges the usefulness of seeking to measure health-system **efficiency**. However, the measurement of efficiency gives rise to a large number of technical problems that have yet to be resolved, as explained in Section XIV. This work requires further development and consultation, and WHO should recognize that it is work-in-progress in any tables it produces.
- XV. SPRG considers that **enhancing policy relevance** is an essential aspect of the HSPA exercise, without which the finest technical endeavours will be irrelevant. WHO has made a number of recommendations for country support and capacity building, all of which appear to offer promise. Their implementation will require careful design and evaluation.

We feel that the independent peer review process has been illuminating and valuable to both WHO and SPRG members, and that the WHO consultation process has already enhanced the effectiveness of the HSPA initiative. We believe that adoption of our recommendations will further enhance the longer-term effectiveness of HSPA, and are pleased to note that many of our comments and suggestions during the review process have already been incorporated into the WHO methodology. More generally, we hope that the usefulness of the peer review process will encourage WHO to embrace the principle of engaging with independent outside expertise on specific HSPA topics, whenever appropriate.

I. THE FRAMEWORK FOR HSPA

1. WHR 2000

WHR 2000 defined the health system to include personal medical care, public health interventions and inter-sectoral actions designed primarily to improve health. It was recognized that the health system contributes towards many outcomes that are socially desirable, including improving health, educational attainment, and individual incomes. WHR 2000 specified a parsimonious set of these goals where the contribution of health actions is sufficiently large to warrant measuring the goals regularly.

A goal was defined as intrinsically valued if raising the level of attainment on that goal is desirable in and of itself. To ensure that each goal measures a different outcome, it was further specified that each intrinsic goal must be at least partially independent of all others, i.e. it is possible to raise the level of attainment of the goal while holding the level of all other intrinsic goals constant. Instrumental goals were defined as outcomes that are desirable because they contribute to attainment of an intrinsic goal.

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 was to improve health, both the average level of population health and its distribution (i.e., to reduce health inequality). 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 also had two components – the average level of and inequality in responsiveness.

The third intrinsic goal was the fairness in household financial contributions to the health system. Although it was recognized that other goals, such as educational attainment and income-earning potential, might meet the criteria of an intrinsic goal, it was judged that it was impractical to measure the impact of the health system on them on a regular basis.

A number of instrumental goals, such as the coverage of health services, were discussed in the text of WHR 2000. But the level of attainment on these goals was not measured in the annex tables. WHR 2000 also defined four basic functions which contribute to intrinsic goal attainment – financing, service provision, resource generation, and stewardship. The text of the

report summarized the available evidence relating these functions to goal attainment and efficiency, but did not define or measure indicators of performance for the functions.

The final concept proposed in the framework was called performance – equivalent to the economic concept of efficiency. It was defined as the system’s contributions to the intrinsic goals taking into account the inputs used to achieve them. The efficiency of systems in producing health and the composite attainment index (made up of the three intrinsic goals) was estimated.

2. Main commentaries and criticisms

There are eight commentaries and four responses that relate to the issue of framework (Almeida et al. 2000; World Health Organization 2001a; DfID 2000; McKee 2001; Murray and Frenk 2001; Navarro 2000; Navarro 2001; Oswaldo Cruz Foundation 2000; Travassos and Buss 2000; World Health Organization 2001b; Williams 2000, 2001; Murray et al. 2001). Only general issues relating to the framework are discussed here. Debates and suggestions on specific indicators, overall attainment, and efficiency are discussed later in this report.

Definition of the health system

Two opposing definitions were expressed. One was that the boundaries of the health system should be drawn tightly around the activities under the direct control of the Ministry of Health – largely personal medical services (WHO Regional Office for the Americas, 2001a). The other extreme was that the boundaries used in WHR 2000 were too narrow, focusing largely on medical interventions and ignoring the broader social and perhaps spiritual components (Navarro 2000; WHO Regional Office for the Americas, 2001b; Ugá et al. 2001; Oswaldo Cruz Foundation 2000).

Goals

Some debate focused on the appropriate intrinsic goals to include for routine monitoring. Again, two opposing views were expressed. The first was that health systems should be concerned with, and judged on, their contributions to population health alone so that population health should be the only intrinsic goal. The opposing view was that not only should health, responsiveness and fairness of financial contributions be included as intrinsic goals, but others should be added as well (WHO Regional Office for the Americas 2001b). The OECD (Hurst 2002), for example, has adopted the concept of responsiveness in its proposed health-system performance framework, but called it “responsiveness and access”. (Access was defined by WHO as an intermediate goal rather than an intrinsic goal – more access

is only valued if it contributes to furthering one of the intrinsic goals). In addition, the OECD has included the level of health financing as an intrinsic goal, although it has not attempted to identify the optimal level for each country.

Other approaches to identifying indicators that could be potentially useful for HSPA include the Essential Public Health Functions (EHPF) of the WHO Regional Office for the Americas. That approach defines 11 key functions involving a mix of inputs, functions and outcomes without a composite index or an explicit statement of which ones contribute more or less to health-system performance. The “benchmarks of fairness” approach of Daniels et al. (2000) is similar. In it, nine benchmarks are used to evaluate the impact of health-system reforms on “fairness” – including assessing the impact on: coverage of key interventions both within and outside the health sector (e.g. literacy, education); barriers and inequalities in access; equitable financing; efficacy, efficiency and quality of care; democratic accountability and empowerment; and autonomy. Each benchmark contains many components and sub-indicators which raters must evaluate subjectively and incorporate into a composite rating for each benchmark on a scale from –5 to +5. No weighting is suggested for possible aggregation across indicators. The benchmarks and their components include indicators that would be labelled intermediate goals in the WHO framework as well as components of WHO’s intrinsic goals (e.g. autonomy is a component of responsiveness in the WHO framework). These approaches focus on very important questions related to the functioning of the health system and, in the case of the benchmarks project, the impact of reforms on fairness. They do not purport to be a comprehensive framework for assessing health-system performance.

Attribution and measurement

All regional consultations pointed out that goal attainment is influenced not only by health actions but by non-health system actions as well. WHO used multivariate statistical analysis to separate the influence of the health system from other possible determinants. Some commentators suggested that it would be useful for decision-making to define and measure indicators of that part of overall goal attainment believed to be determined largely by the activities of the Ministry of Health, either instead of, or in addition to, the outcome indicators defined by WHO. One possible example is the number of deaths due to medical errors which is more directly under the control of the Ministry of Health than all-cause mortality.

Inputs

There was little published criticism of the use of health expenditure per capita as an aggregate indicator of the inputs available to the system, although questions of timing between inputs and outcomes have been raised (Williams 2000; Ministry of Health, Vietnam 2001), which are considered in Section XIV

on Efficiency. Regional consultations also suggested the need to measure inputs to the production of health such as human resources (WHO Regional Office for the Americas 2001b).

Health system functions

Many commentators argued that while information on intrinsic goal attainment was important, it was only a starting point. It was necessary to develop indicators that allowed policy makers to “drill down” so as to discover possible causes of poor performance and ways in which that might be improved (WHO Regional Office for the Americas 2001b; WHO Regional Office for the Eastern Mediterranean 2001). These indicators should be linked to the key function of the health system, which would make the measurement exercise more policy-relevant (Ollila and Koivusalo 2000).

Performance and efficiency – Terminology

In WHR 2000 the term “performance” was used as a synonym for “efficiency” (Williams 2000). At a number of the regional consultations it was suggested that “health system performance assessment” should be defined to include the measurement of goal attainment, as well as the efficiency of input use and the way the system is functioning (WHO Regional Office for the Americas 2001b; WHO Regional Office for Europe 2001), whereas the term “efficiency” should be used in the narrower sense of how well resources are used to produce the desired outcomes.

Focus

Participants in the South-East Asian Regional Consultation suggested that it would be useful to extend the performance assessment exercise to the sub-national level (WHO Regional Office for South-East Asia 2001), while Wibulpolprasert and Tangcharoensathien (2001) argued it could also be used to assess the performance of particular programmes or interventions.

3. WHO responses and proposals

WHO proposes to retain the definition of the health system used in WHR 2000 and to retain the three intrinsic goals. Although WHO recognizes the desire for policy makers to define an appropriate level of health spending, appropriate ways of operationalizing the concept need to be developed. Since publication of WHR 2000, WHO has attempted to make the framework more policy-relevant by defining a set of intermediate indicators that can be of immediate use to policy makers, allowing them to drill down to possible causes of poor performance. They are linked to the four key functions of health systems and are discussed later in this report. It has also begun to

develop ways of assessing the inputs of human resources to the system (see Section III).

4. SPRG comments and recommendations

Definitions

The definition of health systems as proposed by the Secretariat in WHR 2000 is clear and acceptable. The three levels of health attributes, i.e. personal medical, non-personal health services, and inter-sectoral actions, are acceptable and should be used as the 'operational framework'. Given the definition of 'health' in the WHO constitution, which encompasses physical, mental and social well-being, it is suggested that WHO could work with other international agencies to ensure that the impact of health on education and income could be assessed at regular intervals. Some members of SPRG felt that WHO might consider interacting with UNDP to explore if it were possible to modify the Human Development Index (HDI) into something like a Health-Adjusted Human Development Index (HAHDI) by substituting HALE for Life Expectancy in the HDI.

In interacting with other international agencies, it was also suggested that they might do a 'health impact assessment' of their activities on a routine basis.

Goals

The three intrinsic goals – i.e. health, responsiveness and fairness of financial contributions – are operational and acceptable. Countries do care about the level of financing as well, but there is no easy way to operationalize the ideal level of health financing for every country, and inclusion of this might have to be postponed.

SPRG members nonetheless agreed on the importance of retaining measures of the level of financing in future reports (as WHO proposes), and on the benefits of WHO collaborating with any future OECD work on optimal levels of health spending.

Attribution and measurement

SPRG agrees with the Secretariat's proposal to measure the system's contribution to the desired final outcomes. Although this may be a difficult task in developing countries with limited capacity, efforts should be commenced. SPRG also commends the attempt at regular measurement of intermediate goals as proposed by the Secretariat. Data availability and accuracy, scientific soundness of method, including transparency of the

processes, are major concerns. Responses by the WHO Secretariat to these questions have been encouraging in that weaknesses in WHR 2000 have been acknowledged and steps have been taken for improvement.

Inputs

It is reasonable to use health expenditures as the main input. However, data availability in this area is critical and in many countries either there are no reasonable estimates or there are competing estimates. WHO should work with international agencies to standardize methods and estimates on variables such as GDP, expenditures and purchasing-power-parity exchange rates. It should also build capacity in countries in this area. SPRG also commends the Secretariat's attempts to explore the possibility of estimating the quantities of labour and capital stock for all Member States.

Functions

The four main functions are acceptable. SPRG commends the Secretariat's proposal to measure routinely a set of instrumental goals linked to each of the four functions as well as to selected attributes of these functions. It suggests that the work of the benchmarks-for-fairness exercise (Daniels et al. 2000) could provide useful insights. Some members of SPRG suggested that WHO might consider adding one more function, the **organization of health resources** as suggested by Kleczkowski, Roemer, and Van der Werff (1984). This function may be inserted between resource generation and service provision. It is quite logical to think that after we generate resources, we organize them, and then they provide services. However, other members felt that this is really part of each of the other functions.

Performance

SPRG recommends that the term 'performance' should be redefined to include the measurement of goal attainment, as well as the efficiency of input use and the way the system is functioning. 'Efficiency' should then be used more narrowly to represent the concept of value for money.

A strategic plan

SPRG recognizes that the HSPA exercise of WHR 2000 stimulated fresh thinking about health-system performance, and awareness and concerns for better health information (particularly vital registration, health-care financing, morbidity and mortality data, and responsiveness.) It recommends that WHO develop a strategic plan to improve data availability and accuracy of all variables at the global, regional and country levels. Specific plans, including the World Health Survey, should be developed and implemented with clear targets of achievement. Additional resources from funding agencies could also be mobilized for this purpose.

Other suggestions

Issues	Improvements
Measurements	<ul style="list-style-type: none">▪ Request for further disaggregation should be handled with great care. Only very important variables should be measured routinely.▪ Improvement of indicators through more interactive peer review and user feedback.
Data non-availability and accuracy	<ul style="list-style-type: none">▪ Provide more support to sustain national information systems in increasing coverage and accuracy of available data.▪ Improve estimation through more peer review process on scientifically sound techniques.▪ Transparency through detailed explanation of data sources and estimation techniques.
Methodologies	<ul style="list-style-type: none">▪ Demystification through capacity-strengthening, creation of global, regional, sub-regional and country networks/institutes for HSPA. More training, publications, surveys and research are needed.▪ Involve more researchers in future rounds.▪ Use more extensive peer review on methodologies.▪ Description of methods used and commentaries, as well as current improvement strategies, should be presented and open to the public.
League tables	<ul style="list-style-type: none">▪ The decision to continue with the league table is a strategic one, which it is up to WHO to make. If WHO decides to continue publishing league tables, it should include more explanations and precautions. More detail on this issue is discussed in Section XI on 'Composite Indicators'.▪ Incentives may be offered to those at the bottom of the list, e.g., more WHO support, more financial support from donors, reward for improvement in the next round of assessment.

5. References

Almeida, C. M., P. Braveman, M. R. Gold, C. L. Szwarcwald, J. M. Ribeiro, A. Miglionico, J. S. Millar, S. Porto, N. R. Costa, V. O. Rubio, M. Segall, B. Starfield, C. Travassos, A. Ugá, J. Valente, and F. Viacava (2001): Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357(9269): 1692-1697.

Daniels, N., J. Bryant, R. A. Castano, O. G. Dantes, J. S. Khas, and S. Pannarunthai (2000): Benchmarks of fairness for health care reform: A policy tool for developing countries. *Bulletin of the World Health Organization*, 78(6): 740-750.

DfID (Department for International Development, UK Health Systems Resource Center) (2000): World Health Report 2000: Summary and comments. Draft 10 July 2000.

Hurst, J. (2002): Performance measurement and improvement in OECD health systems: Overview of issues and challenges. *Measuring Up: Improving Health System Performance in OECD Countries*. Organization for Economic Co-operation and Development.

Kleczkowski, B. M., M. I. Roemer, and A. Van der Werff (1984): National Health Systems and their reorientation towards health for all guidelines for policy-making. *Public Health Papers*, No. 77. Geneva, Switzerland: World Health Organization. [http://whqlibdoc.who.int/php/WHO_PHP_77.pdf]

McKee, M. (2001): Measuring the efficiency of health systems. *British Medical Journal*, 323(7308): 295-296.

Ministry of Health, Vietnam (2001): Comments and suggestions of Vietnam Ministry of Health/Health Policy Unit as regards the World Health Report 2000.

Murray, C. J. L. and J. Frenk (2001): World Health Report 2000: A step towards evidence-based health policy. *Lancet*, 357(9269): 1698-1700.

Murray, C. J. L., J. Frenk, D. Evans, K. Kawabata, A. Lopez, and O. Adams (2001): Science or marketing at WHO? A response to Williams. *Health Economics*, 10(4): 277-282.

Navarro, V. (2000): Assessment of the World Health Report 2000. *Lancet*, 356(9241): 1598-1601.

Navarro, V. (2001): World Health Report 2000: Responses to Murray and Frenk. *Lancet*, 357(9269): 1701-1702.

Ollila, E. and M. Koivusalo (2000): Values, ideologies and evidence-based recommendations -- The World Health Report 2000: WHO's health policy drifting off course. In eds. U. Häkkinen and E. Ollila. *The World Health Report 2000: What does it tell us about health systems? Analyses by Finnish Experts*. Helsinki: National Research and Development Center for Welfare and Health (STAKES).
[<http://www.stakes.fi/english/publicati/Publications.htm>]

Oswaldo Cruz Foundation, Brazilian Ministry of Health (2000): Report of the workshop "Health Systems Performance: The World Health Report 2000". 14-12-2000, Rio de Janeiro.

Travassos, C. and M. Buss (2000): The controversial World Health Organization report. Editorial. *Cadernos de Saude Publica*, 16(4): 890-891.

Ugá, A. D., C. M. Almeida, C. L. Szwarcwald, C. Travassos, F. Viacava, J. M. Ribeiro, N. R. Costa, P. M. Buss, and S. Porto (2001): Considerations on methodology used in the World Health Organization 2000 Report. *Cadernos de Saude Publica*, 17(3): 705-712.

Wibulpolprasert, S. and V. Tangcharoensathien (2001): Health systems performance: what's next? *Bulletin of the World Health Organization*, 79(6): 489.

Williams, A. (2000): Science or marketing at WHO? A Commentary on 'World Health 2000'. *Health Economics*, 10(2): 93-100.

Williams, A. (2001): Science or marketing at WHO? Rejoinder from Alan Williams. *Health Economics*, 10(4): 283-285.

World Health Organization (2001a): The methods and data used in the World Health Report 2000: A response to Almeida et al. Geneva, Switzerland: World Health Organization.

World Health Organization (2001b): The methods and data used in the World Health Report 2000: A response to the commentary made by the Brazilian delegation to the Executive Board, 17 and 19 January 2001.

WHO Regional Office for the Americas (2001a): Regional consultation of the Americas on Health Systems Performance Assessment, Background Document: Critical issues in health systems performance assessment. Washington, D.C.

WHO Regional Office for the Americas (2001b): Regional consultation of the Americas on Health Systems Performance Assessment: Final Report. 8-5-2001, Washington, D.C.

REPORT OF SPRG ON HSPA

WHO Regional Office for the Eastern Mediterranean (2001): Report on the Regional Consultation on the Conceptual Framework for Health System Performance Assessment. 9-7-2001, Ain Saadeh, Lebanon.

WHO Regional Office for Europe (2001): Report of the Regional Consultation on Health Systems Performance Assessment. 3-9-2001, Copenhagen.

WHO Regional Office for South-East Asia (2001): Report of the regional consultation and technical workshop on health systems performance assessment. 18-6-2001, New Delhi, India.

II. HEALTH SYSTEM INPUTS

Health system inputs include both physical and financial resources. The concepts and measurements related to these inputs play two distinct roles in WHR 2000. First, they are important to estimating the efficiency of the health system. WHR 2000 explicitly utilized total health spending as the aggregate input in its efficiency estimates, but spending was implicitly being used to reflect the application of physical resources toward improving health.

Secondly, health system inputs were discussed in terms of health system functions. In the case of physical inputs, the conceptual framework discussed both their supply (resource generation) and utilization (under service provision). This section focuses specifically on the WHR 2000 report of financial inputs through the use of National Health Accounts.

1. WHR 2000

Concepts

National Health Accounts (NHA) are a method for quantifying the financial flows of the health system in a comprehensive, consistent, and integrated manner. In WHR 2000 and WHR 2001, the estimates of health expenditure for 191 countries are reported and disaggregated by source. The principal categories are public and private spending, which are further disaggregated into tax-funded, social insurance, out-of-pocket, and private insurance expenditures. WHR 2000 presented the first global NHA estimates using data for 1997. These data have been revised using additional statistical sources. The revised data for 1997 and new estimates for 1998 are presented in WHR 2001. WHR 2001 also included a new sub-category: externally-financed health spending.

Methods

The health system is quite complex, but the nominal value (and, where unavailable, the imputed value) of resources funded and spent in a health system have to be equal to the sum of the value of all goods and services delivered. Using this identity to enforce consistency, NHAs organize health financing information by selecting a group of dimensions that are useful for analyzing this complexity. These dimensions are then summarized in a series of matrices that provide information on expenditures by source, financing agents, providers, and uses. By using common classifications, it becomes possible to learn through comparisons across countries and over time. Efforts are made to report expenditures on an accrual basis (i.e., when the resources are consumed not when payments are made) wherever possible (Poullier and Hernandez 2000).

Data/Evidence/Sources

Complete data were not available for all countries. For WHR 2000, the information was based on 67 country NHA reports for various years. Of these, 30 traced expenditures through the main components of health-care financing: resource mobilization, resource allocation, and service and goods provision. Estimates had to rely on partial information for 124 countries, for which various sources of information on health expenditure were consulted. Figures presented in the WHR are in US dollars at official exchange rates, and also in international dollars converted at purchasing power parities (PPP).

2. Main commentaries and criticisms

Concepts

No one questioned the usefulness of NHA for estimating health expenditure per capita as an aggregate indicator of the inputs available to the system. Nevertheless, some comments were made on the basic concepts, methods and data.

Some reviewers were concerned that WHR 2000 assigned too much importance to the data reported in the National Health Accounts, particularly as the single most important input in calculating the efficiency of health systems. One article stated that "*NHAs seem to be treated as a sort of panacea ... for the purpose of restoring productivity in the Member States' health systems*" (DfID 2000).

Another criticism was that NHA are not fully consistent with the WHO definition of the health system (McKee 2001). The reported NHA data concentrate mainly on personal medical and non-personal health-services expenditure. Intersectoral actions and the production of resources, which are emphasized in WHR 2000 as integral parts of the health system, are not always included. The figure for health expenditure on all the activities of intersectoral actions in promoting health "*is nowhere to be found in any national health accounts*" (McKee 2001). Therefore, using NHA data to measure health-system resources is misleading since it forces a comparison of outcomes with health *services* inputs rather than with health *system* inputs.

Methods

A range of criticism focused on the non-comparability of expenditure categories and definitions. Different agencies classify expenditures differently, and do not have the same definition for functions and services (Ministry of Health, Lebanon 2000). Comparability is difficult because of diverse national standards, in addition to differing concepts of boundaries, dimensions, and classification systems. Furthermore, standardized regional

reporting systems are lacking, and it is very difficult to achieve consensus by policy makers regarding the framework and content of NHA (WHO Regional Office for South-East Asia 2001).

Due to the important role played by these figures, and the questions regarding methodology and data sources, it was argued that future reporting needs to discuss explicitly levels of uncertainty for expenditure estimates. This is highlighted by the debates around WHR 2000 on the assessment of outcomes, inputs and efficiency with widely varying uncertainty intervals.

Data/Evidence/Sources

There is lack of comparability between the different data sets. Collecting information on the distribution of expenditures by function, and linking expenditures with utilization is difficult given the state of most countries' health financing data. The NHA estimates were questioned on the basis that the quality, validity and reliability of the data available in the countries is variable and frequently poor. They noted that discrepancies exist between expenditure data from different sources and questioned the process of reconciling these varying estimates.

Sources used to estimate health expenditures are not always complete. Sometimes they provide data only on the public sector while others concentrate on private expenditures. There is double-counting as well as gaps in coverage. Estimations of private spending were a particular subject of such criticism. The quality and reliability of such estimates remains uncertain because of incomplete coverage, unrepresentative surveys, and the likelihood of double counting expenditures (WHO Regional Office for South-East Asia 2001). For example, some commentators argued that reliable estimates of out-of-pocket expenditure were unavailable in up to 75% of countries (WHO Regional Office for South-East Asia 2001). Development of better methods to estimate private spending were recommended.

Some countries disagreed with the estimates of health expenditure reported in WHR 2000, on account of the data being outdated, the source of data being insufficient, and methods applied.

Some countries also disagreed with estimates of exchange rates (official and/or PPP), and the sources of PPP exchange rates used to convert local currency units into international dollars were questioned.

Countries argued that WHO had a responsibility to provide necessary technical assistance to prepare or improve NHA data. They argued that WHO's data collection efforts should more closely integrate with capacity-building in the countries. Regional activities may be needed to ground NHA in countries and to build better evidence for policy.

Specific problems were cited in WHO's estimation of external resources applied to health in WHR 2001. It was noted that there are large discrepancies between NHA surveys and data on public expenditure on health reported in national and international publications ("The Health Dime"). The problem cannot be ignored because in some cases external resources account for a third or more of total resources spent on health. Some of the limitations include:

- Disadvantages of the existing registration system: data reported under the IMF classification of expenditure by function appear in many countries to be institutional rather than functional, and omit expenditure on health-enhancing interventions by ministries other than the Ministry of Health. They frequently report planned rather than actual outlays.
- The multiple channels through which external resources flow into a country are not always reported.
- There are problems in tracing and valuing in-kind flows. For instance, donations of equipment, stocks of vaccines, and medicines are not always reported or given a value.
- Private grants and loans are not registered.

Overall, the importance of explaining the findings and checking their validity was emphasized.

3. WHO responses and proposals

Concepts

WHO maintained that the concept used in NHA is close to the definition of the health system and argued that much of what gets reported depends on the availability of data. WHO recognizes that statistical imperfections exist and in order to address the criticism of inconsistencies in the definition of the health system, the new framework for NHA incorporates further breakdowns of expenditure.¹

WHR 2001 introduced two additional categories of expenditures, viz. external resources for health as a percentage of public expenditure on health, and private insurance and other pooled expenditure. A further enlarged dataset is anticipated for WHR 2002 which will include a broader time-frame (1995-1999). Also, trends for 1970-2000 are in the process of estimation (with completion planned after 2002). WHO also proposes the introduction of further breakdowns of expenditure to the extent possible:

- (i) by type of function or provider (inpatient, outpatient, long-term care);
- (ii) by type of resource (capital, labour, consumables);

¹ These will appear in the NHA *Producer's Guide* currently under preparation. It is co-funded by the World Bank, WHO and USAID, and is being jointly prepared by those agencies and a team from the Harvard School of Public Health.

(iii) by type of provider.

In order to improve cross-country consistency and provide technical assistance, WHO is in the process of preparing a *Producer's Guide* to offer a common framework for NHA. WHO has also initiated a capacity-building programme at regional and national levels (Africa and Eastern Europe) with other activities also to respond to demands from the Americas, the East Mediterranean, and Asia. WHO is collaborating with other international organizations on developing the methodology for NHAs and on technical training.

Methods

A prototype NHA has been initiated because of large gaps in data (particularly for the private sector), non-availability of required disaggregations, cash- and not accrual-data. Boundaries reflect different degrees of the private-public mix ("in search of commonality").

Data/Evidence/Sources

WHO has redoubled its efforts to interact with national authorities and other international agencies in a continuing process of updating estimates. WHO has specifically initiated a consultation process with countries in order to validate the information compiled. For example, estimates for 1998 and the revisions for 1997 were sent to countries for comment, and the changes agreed with them were reported in WHR 2001.

Since WHR 2000, WHO is working on a method to improve the calculation of purchasing power parities, and is collaborating with the World Bank for the next round of the International Comparisons Project.

To deal with the problems identified in estimating external resources, WHO suggests that:

- Data should be collected from both the external agencies and the recipient countries (through the questionnaire listing input categories using the major functional classifications);
- The value of in-kind transfers should be calculated at replacement cost;
- The resources allocated only by the external agency directly to the population and health institutions – as cash or in-kind – should be counted.

4. SPRG comments and recommendations

In the new framework of the NHA, WHO intends to incorporate the main indicators concerned with inputs examination for the purposes of HSPA. This

is evident from an examination of the modified template for NHA, which is closer to the classifications in the *Producer's Guide*. The purpose is to document (through the NHA) the problem of existing imbalances between different types of resources. However, NHA provide only one-year expenditures, whereas for the purposes of investment planning the stock of available capital is needed – which incorporates past investment decisions too. Preparing a time-series of NHA will indicate the changing demands and supplies of the health system over time.

NHA are based on accountancy principles which do not readily offer estimates of data uncertainty. This may conflict with the statistical methods used in other aspects of HSPA. There is a problem with the availability and accuracy of the data being presented. No explanation is available on the methods used to estimate missing data (when the nominal values are not available). There is little or no information on the approach that is applied to choose among different values reported by different sources, and to decide on the level of expenditure when no information exists (e.g., expert opinion, imputed from international patterns, some sort of average, etc). WHO should consider developing methodologies for indicating uncertainty in the financial data.

WHO is attempting to enter into a dialogue with countries for continuous improvement of NHA. However, more transparency is needed. WHO is consulting countries to establish whether the series are plausible and to fill gaps. Basic macro-variables – such as general government total expenditure – should be reviewed, and information to fill missing data should be obtained at the national level. WHO has to specify the procedures for examining the accuracy in the NHA data.

It is extremely important that WHO work towards the standardization of classifications. The classifications applied are usually country-specific, adopting the OECD International Classification of Health Expenditures' system to the country's own situation. Also, the number and scope of breakdowns presented are different, depending on the data available.

Apart from standardizing classifications, the definitions of categories have to be uniform. Different items that are similarly labelled may be included in health-care costs. SPRG recommends that in order to achieve uniformity among expenditures labelled identically and to ensure comparability across countries, WHO should clarify the content of each category incorporated into the NHA template. Explicit definitions are needed. Clarification is also needed to ensure the consistency of NHA with the definition of the health system, so as to be able to identify the inputs to the health system from which the outcomes are evaluated.

The weak basis of PPP estimates can be highly influential on a country's measured performance. However, it remains a preferable alternative to

conversion using official exchange rates. We would encourage WHO to continue actively exploring improved PPP estimates.

Additional household surveys are often undertaken to elicit private out-of-pocket expenditures. Unfortunately, attention is not always paid to capital investment in private facilities, and to the health expenditure incurred by a country's citizens when they are abroad. Another problem is health spending by citizens of neighboring countries (especially living in border areas) who work and pay health insurance premiums in one country but live and use health services in another. WHO should determine whether such spending should be included.

In terms of basic accounting principles, income and expenditure must be balanced in the NHA. The value of resources funded and spent in a health system should be equal to the sum of the value of all goods and services delivered (Poullier, Hernandez, and Kawabata 2001). However, expenditures (consumption and investment) are not necessarily equal to the sum of the value of goods and services provided. The amount of investment in resource generation (for training and construction) is usually different from the cost of factors employed, for the following reasons:

- migration and unemployment of human resources in the health system;
- unfinished construction;
- buildings are not depreciated (no consumption of fixed capital is taken into account) until they are finished.

WHO should examine further the most appropriate treatments of these complications.

There is an extremely important role for WHO to play in *capacity-building* of countries.

- We propose that WHO publish the *Producer's Guide* as soon as possible, in order to increase capacity-strengthening in countries;
- WHO should work for greater harmonization across international agencies;
- WHO should strengthen support to the NHA regional networks and find ways of improving interaction among users of NHA. This will also help in capacity-building.
- For countries with limited capacity, the measurement process for health expenditure flows should be incorporated into the regular UN System of National Accounts.

5. References

Anonymous (2000): *The Health Dime*, No. 1. Geneva, Switzerland.

DfID (Department for International Development, UK Health Systems Resource Center) (2000): World Health Report 2000: Summary and comments. Draft 10 July 2000.

Dorabawila, T., S. De Silva, J. Mendis, and E. Rasell (2001): WHO Fairness in Financing Study: Estimates for Sri Lanka 1995/96 using WHO methodology. Colombo, Sri Lanka: Institute of Policy Studies of Sri Lanka.

McKee, M. (2001): Measuring the efficiency of health systems. *British Medical Journal*, 323(7308): 295-296.

Meerding, W. J., L. Bonneux, J. J. Polder, M. A. Koopmanschap, and P. J. Van der Maas (1998): Demographic and epidemiological determinants of health care costs in Netherlands: Cost of illness study. *British Medical Journal*, 317: 111-115.

Ministry of Health and Institute of Policy Studies (2001): Sri Lanka National Health Accounts: Sri Lanka National Health Expenditures 1990-1999. Colombo, Sri Lanka: Institute of Policy Studies of Sri Lanka.

Ministry of Health, Lebanon (2000): Lebanon National Health Accounts: Executive Summary (Working Draft).

Poullier, J. P. and P. Hernandez (2000): Estimates of National Health Accounts for 1997. Global Programme on Evidence for Health Policy Discussion Papers, No. 27. Geneva, Switzerland: World Health Organization.

Poullier, J. P., P. Hernandez, and K. Kawabata (2001): National Health Accounts: Concepts, data sources and methodology. Global Programme on Evidence for Health Policy Discussion Papers, No. 47. Geneva, Switzerland: World Health Organization.

Poullier, J. P., P. Hernandez, K. Kawabata, W. Savedoff, C. Indikadahena, and R. Zeramdini (2001): Global and national spending patterns: Results from National Health Accounts.

WHO Regional Office for South-East Asia (2001): Report of the regional consultation and technical workshop on health systems performance assessment. 18-6-2001, New Delhi, India.

III. RESOURCE GENERATION FUNCTION

1. WHR 2000

In WHR 2000 the available evidence on the links between the resource generation function and health-system performance was summarized. WHO argued that whatever the level of inputs, there was an efficient way to combine them. Significant imbalances between different types of productive resources existed in many settings, and countries must address a number of complex questions such as:

- What is the most cost-effective balance between different types of productive resources and how can this be achieved?
- How much effort should be devoted to developing new resources (e.g., investment) compared with developing strategies and incentives to improve the use of existing resources?

No attempt was made to define or measure indicators of how the resource generation function was being performed. Health expenditure per capita was the only source of information on health-system inputs used for performance assessment.

2. Main commentaries and criticisms

There were few comments and criticisms of the WHO approach to this function apart from the general comment that WHO needed to develop the links between each of the four key functions and the performance of the system as a whole (WHO Regional Office for the Americas 2001).

Human resources were seen as particularly important because health systems are labour-intensive and expenditure on personnel is usually the largest single item of recurrent health expenditure. Health systems require not only a sufficient number of qualified and experienced staff to function well, but an appropriate mix between the different types of human resources. Changing the mix will not, however, solve all problems and some commentators (e.g., DfID 2000) suggested that special attention should also be paid to the following issues:

- the difficulty of reorienting staff from one activity to another
- the problem of low productivity of human resources which was seen to be linked closely to the issue of remuneration;

- development of ways to measure and improve the quality of human resources, perhaps linked to realistic estimates of the level of outputs that the various inputs might be expected to deliver (DfID 2000).

Similar comments were made in relation to physical capital where deficits in the stock of assets (e.g., buildings and equipment) can be a real constraint to the delivery of effective interventions. 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 for staff costs or the maintenance of the physical capital stock, which results in efficiency losses. Goal attainment is a function of the number and type of health facilities and equipment available. Various types of incentives and legislation influence how capital is purchased, used and maintained. Hence it is not just a matter of counting the availability of resources, but of ensuring that the mix is appropriate and the resources are used efficiently (Anell and Willis 2000).

3. WHO responses and proposals

WHO has proposed a set of indicators for each of the four key functions, which will help decision makers identify practical areas where performance can be improved. For the resource-generation function, WHO proposes to focus on investment in the production of resources and maintenance of their quality and productivity. The management and deployment of resources will be assessed under the service-provision function.

Generation of human resources for health

The following indicators are proposed:

- Total annual investment in human resources (HR) as a percentage of total health expenditure;
- The number of new entrants to educational institutions that train health-care professionals divided by the total stock of health-care personnel;
- The total stock, composition and distribution of human resources for health;
- Migration of human resources.

As an input to this exercise, it will be necessary to explore the feasibility of estimating the quantity of different types of labour inputs currently available to the health systems of Member States. To do this, WHO will develop a global database on human resources. Data will be collected on the quantity and characteristics of different provider groups, partly through the World Health Survey (for which a draft survey module has been developed).

WHO also proposes to develop a human resources policy. The purpose will be to synthesize the evidence on the effects of different human resources policies on the performance of health systems. This will allow different parts

of WHO to work together to develop a coherent set of strategies that can help Member States to improve the performance of this function. The Organization acknowledges that there is a strong need for capacity building in countries to achieve effective policy-making in this area.

Physical resource generation

The following activities are proposed by WHO:

- (i) To explore the feasibility of estimating the quantity of different types of capital stock (e.g., health facilities, equipment).
- (ii) To develop and apply methods for measuring the physical capital stock available to the health system (e.g., the value of buildings and equipment).
- (iii) To monitor:
 - Annual new investments in health facilities as a percentage of total health expenditure;
 - Annual expenditure on maintenance as a percentage of annual investment in health facilities;
 - The total stock of facilities (current value) as a proportion of GDP.

Pharmaceuticals and medical devices

WHO proposes to measure investment in medical devices as a component of its work on health facilities. Measuring the availability and utilization of essential drugs and other consumables is included in WHO proposals to define indicators of the health services provision function (Section IV).

Knowledge

WHO proposes to measure total annual investment in health research and development. WHO has also started a process to develop the performance assessment of health research systems, and plans to publish its findings in WHR 2004.

4. SPRG comments and recommendations

Human resource generation

SPRG supports the need to develop indicators for each of the four functions of the health system, including resource generation. *Inter alia*, this will help to generate evidence about the influence of the composition of human resources on the attainment of health-system goals.

Among the problems related to human resources, the lack of standardization in definitions of human resource categories needs to be addressed. More

attention also needs to be paid to non-medical professionals and the migration of human resources.

The methodology proposed by WHO for estimating National Health Accounts (NHA) incorporates an additional category of expenditure – investment for human resource generation (production and continuing development). SPRG believes that 'annual investment in human resource generation as a percentage of total health expenditure', one of the indicators proposed by the Secretariat, is too general as an indicator of efficiency for the human resource generation function. The inclusion of maintenance costs for human resources is also desirable in an assessment of efficiency, even though it may be difficult to collect these data. WHO should explore whether maintenance costs could be included as part of NHA.

Concerning the breakdown of human resources categories, SPRG questions whether the six provider categories proposed by WHO are sufficient. Other categories such as public-health physicians, preventive-care professionals and traditional health-care providers might also be important. On the other hand, SPRG recognizes that too many categories may overburden health-information systems and make data collection difficult.

Owing to the functional substitution between different categories of human resources, which often occurs in resource-poor areas, data on the quantity and characteristics of selected categories may not represent the functional profile of human resources in certain areas.

There is concern about the apparent tendency of WHO to follow the trend of focusing on curative care that is observable in many countries. It is recommended that WHO should pay more attention to traditional public-health occupations in its work on human resources.

SPRG recommends that WHO reviews its work on the migration of human resources with a view to developing an indicator that takes into account the dynamic character of the process.

Indicators to assess performance of human-resource generation should follow the general framework of HSPA, i.e., they should include the quantities of resources available, their distribution, and their efficiency.

Members of SPRG expressed the need for a parsimonious set of indicators related to shortage (demand minus supply), equity (distribution), and efficiency of human resource generation. One possible approach may be summarized in the two tables below.

III. Resource Generation Function

Table 1. Matrix for the assessment of human resource generation

Selected categories	Level			Equity		Efficiency of production
	Adequacy (number, density)	Skill mix	Quality	Fairness of finance of HR production	Distribution of new entrants	
Doctors Nurses Midwives, Public-health workers, Dentists Pharmacists, Managers, Traditional health workers Etc.					Composition (social, demographic, income)	Per capita investment (Investment per trained person)

Table 2. Matrix for the assessment of human resource maintenance and utilization

Selected categories	Level		Equity	Productivity
	Remuneration	Incentives	Distribution	
Doctors Nurses Midwives, Public-health workers, Dentists, Pharmacists, Managers, Traditional health workers, etc.	Possible indicators: - Range, - Timelines (are the salaries paid on time, regularity of payments), - Adequacy (e.g., in comparison with other countries in the region or countries with similar national income)	Non-pecuniary	Distribution among socio-demographic groups	Possible indicators: - No. of Full-Time Equivalents (FTE) per bed occupancy in hospital, or - No. of FTE per visit

These suggestions are made for consideration and further development by WHO.

Physical resources

Investment decisions have an impact on the type of services provided and the geographical distribution of the services. The health system needs to

take account of the current condition of the health-care facilities infrastructure, i.e., the physical capital stock.

For operational efficiency, no standards exist either on the proportion of total health expenditure that should be devoted to investment in physical infrastructure, or on the ratio of maintenance and operating costs to investment.

SPRG welcomes the approach proposed by WHO to establish a core set of equipment to be measured, which can be used to assess resource availability, and to test the feasibility of collecting such information in demonstration countries.

Specific comments on data for indicators:

(i) What to collect. For estimation of the current value of physical inputs, a standard procedure needs to be applied so as to assure comparability across countries. In the first instance, WHO should collect data only on the number and type of selected facilities, equipment, etc., their anticipated physical depreciation, and their distribution in each country. Subsequently, appropriate modelling needs to be undertaken for the imputation of values.

(ii) Sources of data. In general, no agency collects data on the number and types of all fixed assets. Central and local governments often have statistics on specific equipment, e.g., MRI and CT scans. WHO should collect the necessary data but avoid duplication with other data-collection bodies. Financial reports to statistical authorities are a cheaper source of data than direct measurement strategies, although they may be partial. Given different arrangements in countries concerning ownership and management of buildings, information on public facilities may be available only in separate reports (provider reports show only maintenance and operating costs; local government reports include the value of buildings and their depreciation). In consequence, additional surveys may have to be undertaken.

Pharmaceuticals and medical devices

SPRG supports WHO proposals for including pharmaceuticals and medical devices as important resources to be measured.

Knowledge

SPRG commends the WHO initiative to measure the performance of national health-research systems and publish its findings in WHR 2004. The lessons learned from HSPA will be useful in this respect.

5. References

Anell, A. and M. Willis (2000): International comparison of health care systems using resource profiles. *Bulletin of the World Health Organization*, 78(6): 770-778.

DfID (Department for International Development, UK Health Systems Resource Center) (2000): World Health Report 2000: Summary and comments. Draft 10 July 2000.

WHO Regional Office for the Americas (2001): Regional consultation of the Americas on Health Systems Performance Assessment: Final Report. 8-5-2001, Washington, D.C.

IV. SERVICE PROVISION FUNCTION

1. WHR 2000

WHR 2000 described four functions of the health system – financing, resource generation, service provision, and stewardship – and summarized the available evidence about their links to outcomes and health-system performance. In the text of the Report, service provision was defined as the way inputs are combined to allow the delivery of a series of interventions or health actions.

Three main aspects of health-service provision were identified:

- Priority setting – choosing the appropriate mix of interventions;
- Organization of service delivery – choosing the appropriate level for delivering interventions and the degree of integration;
- Aligning provider incentives to ensure that performance is optimized.

Coverage was seen as an intermediate goal, something that was valuable because it contributed to the intrinsic goals. No attempt was made to define and measure indicators of how this function was being performed, or to assess the coverage of key interventions.

2. Main commentaries and criticisms

Assessing service provision attracts attention because it is directly related to the daily management of the health system, and impacts are immediate and visible. WHR 2000 focused mainly on the measurement of intrinsic goals. As with the other functions, Member States and policy makers expressed a desire for practical applications of the assessment exercise. During the regional and technical consultations on HSPA, the development of instrumental goals has been consistently emphasized as a way of allowing policy makers to 'drill down' to find practical ways of improving system performance. At all regional consultations coverage was recognized as one of the key intermediate goals that should be routinely monitored.

3. WHO responses and proposals

WHO proposed three focus areas for health-service provision: (i) health-system inputs, (ii) organizational structure and processes, and (iii) the

quantity and quality of personal and non-personal health services in relation to the health-care needs of the population (Adams et al. 2000).

Nine domains are proposed for these areas in order to assess and monitor the management and development of the health system. Health-system inputs are measured through: (i) recurrent costs of service provision; (ii) physical availability of inputs; (iii) skill mix of health-care personnel; and (iv) utilization of medical equipment and structures. The organizational structure of the system and the process of health-service delivery are assessed through: (i) the level and type of autonomy and integration; and (ii) incentive structures. The outcomes of the service-provision function will be reflected in the intrinsic goals of health and responsiveness, both in terms of overall level and distribution. Two concepts – effective coverage and provider performance – are proposed as instrumental goals.

The concept of effective coverage was developed at a technical consultation in Brazil in August 2001. WHO subsequently proposed that it should incorporate the traditional concepts of access, utilization, and effectiveness (Shengelia et al. 2001). Coverage is an integrated concept using these three traditional elements, and is defined as the probability of receiving a necessary health intervention conditional on the presence of a certain health problem or health-care need. WHO further proposes five domains of coverage – availability, accessibility, affordability, acceptability and effective coverage.

Effective coverage can be estimated at the population level or at the individual level. WHO recommends estimation at the individual level, allowing the estimation of inter-individual inequality in (the probability of) coverage. Effective coverage at individual level can be measured by five steps using data on: prevalence and incidence of diseases; occurrence of interventions in the population; individual observable and unobservable characteristics; health-system characteristics; and effectiveness of intervention (Murray et al. 2001). The Secretariat further proposes that reducing inequality of effective coverage should be an instrumental goal, measured using methods developed previously.

Eight areas and 19 indicators are proposed for regular measurement using the criteria suggested by experts at the technical consultation.

4. SPRG comments and recommendations

The conceptual frame that WHO proposes for considering the service-provision function, consisting of three focus areas and nine measurement domains, seems to be useful. But the relationship between personal and non-personal health services needs further development, particularly in respect of the instrumental goals proposed by WHO for this function, i.e.,

effective coverage and provider performance. In the domain for management of service provision, management-oriented concepts such as autonomy, integration and incentives should be more clearly delineated.

Because targets for intermediate goals of the health system are more manageable in the short term than targets for intrinsic goals, outcome-related process indicators such as effective coverage will be very useful for policy makers and field workers. The assembly of a parsimonious set of indicators of the intermediate goals is an essential step towards enhancing the policy relevance of HSPA.

SPRG agrees that it is highly desirable for WHO to develop indicators of service provision. The categories of inputs, organizational structure, and health services appear to be conceptually sound. However, considerable further work is needed to develop operational indicators. WHO should develop a set of criteria for evaluating such indicators. The indicators should be clear, appropriate, understandable, measurable, and where necessary country-specific.

The process of indicator selection is important, and must involve relevant specialists and field workers. In presenting these indicators, their relationship to other functions such as resource generation – and to the intrinsic goals of health and responsiveness – has to be spelt out. To enhance policy relevance, it would be useful in some settings to measure these indicators at the sub-national level or even at the level of health institutions.

There has recently been a worldwide concern for improvement in the quality of health care, and in several countries quality of care has been redefined to include patient safety (IOM/NAS 2001). SPRG endorses the proposal to develop the notion of provider performance as an instrumental goal, including the concepts of quality and safety.

The analytical framework being developed by WHO for the other instrumental goal of effective coverage holds great promise, but needs to be exposed to detailed peer scrutiny and to incorporate feedback from external experts.

The concept of effective coverage can be important in quantifying the gap between efficacy and effectiveness of many interventions. However, the way that the proposed components of coverage – availability, accessibility, affordability, acceptability and effective coverage – relate to the more traditional concepts of access, utilization and effectiveness, needs to be explained clearly to policy makers.

SPRG endorses the development of carefully chosen measures of coverage that can be shown from research evidence to be linked to the achievement of the intrinsic goals. The use of such indicators is an important step in addressing the difficulty that some outcome indicators relate not only to the current period but reflect the results of health-system activity in the past.

The choice of the type of interventions that are routinely monitored for coverage should be guided by the criterion that these interventions are expected to be significant determinants of population health. (HALE).

The development of indicators of coverage not yet linked through research to the achievement of the intrinsic goals should be approached with caution. Use of such indicators may encourage some nations to introduce interventions that are subsequently shown to be ineffective. We suggest instead that appropriate research be commissioned to identify effective interventions.

Some care is needed in the presentation of coverage data. Interventions that are cost-effective in some countries may not be so in others, and crude rankings will be inappropriate.

Measuring inequality in effective coverage is useful because it is directly amenable to policy and is a determinant of inequality in health outcomes. But the method of measurement should be carefully developed and different alternatives explored.

Finally, in keeping with the general approach of WHO to examine both inequality and deprivation in the intrinsic goals (e.g., fairness in household financial contributions), SPRG recommends that deprivation in coverage should be measured alongside inequality. This will require specifying a minimum threshold level of coverage and estimating the percentage of individuals who fall below it. Identifying individuals with a low probability of coverage of key interventions would be very useful for policy purposes.

5. References

Adams, O., B. Shengelia, B. Stilwell, I. Larizgoitia, A. Issakov, S. Kwankam, and F. Jam (2000): Provision of personal and non-personal health services: Proposal for monitoring. Global Programme on Evidence for Health Policy Discussion Papers, No. 25. Geneva, Switzerland: World Health Organization.

IOM/NAS (2001): Crossing the Quality Chasm: A New Health System for the 21st Century.

Murray, C. J. L., B. Shengelia, N. Gupta, and O. Adams (2001): Inequality in coverage: concepts and measurement strategies. Global Programme on Evidence for Health Policy, mimeo. Geneva, Switzerland: World Health Organization.

Shengelia, B., O. Adams, M. Thieren, Y. Berchmans, Y. Kwankam, and C. J. L. Murray (2001): Measuring the coverage of critical interventions through household surveys. Global Programme on Evidence for Health Policy, mimeo. Geneva, Switzerland: World Health Organization.

V. FINANCING FUNCTION

1. WHR 2000

WHR 2000 (Chapter 5) analyzed health financing as one of the four principal functions of health systems. It categorized financing into the collection of funds, pooling of resources, and purchasing of services. It highlighted the advantages of health financing mechanisms that collect resources from a wide base, pool risk between the sick and the healthy and between rich and poor, and that allocate resources and purchase services strategically.

2. Main commentaries and criticisms

Overall, there was almost no direct criticism of the financing function part of WHR 2000. Nevertheless, references in some of the broader critiques of the report drew on several aspects of the financing function chapter.

The main criticisms relating to the chapter on the financing function were that the analysis was ideologically driven and not based on evidence. Some commentators viewed the framework as inherently biased towards increasing private sector involvement in insurance and health financing (Almeida et al. 2001; Oswaldo Cruz Foundation 2000; Navarro 2000; Navarro 2001a; Navarro 2001b; Häkkinen and Ollila 2000; Van der Stuyft and Unger 2000). Such critiques noted the attention given to the analytical separation of financing and purchasing, the high fairness-in-financing ranking of certain countries (such as Colombia) that have engaged in market-oriented reforms, as well as discussions of a role for private provision. These papers argued that the Report ignored evidence regarding problems with managed competition, private insurance, and other kinds of market-oriented reforms.

An opposing view was expressed by Helms (2000) who perceived the health financing approach taken in WHR 2000 to be inherently biased against private sector involvement. The author argued that if countries followed this approach to health financing they would dull incentives for progress in medical technology and health-service provision.

The response to these arguments of ideological bias towards either "market orientation" or "central planning" was summed up by Murray and Frenk (2001). They argue that the WHR was not advocating any particular policy stance, but rather calling for more systematic evidence in how health systems affect the final goals. According to them, WHR 2000 states "... there is no evidence that systems relying a great deal on public funding will necessarily be more efficient than systems with a greater degree of private sector

involvement, or vice versa. Whether this is seen as a Marxist or capitalist conclusion depends entirely on the ideology of the commentator and the motivations for their commentaries. We see it simply as a summary of the best available evidence at present.”

Another criticism was that the Report did not link the analysis of the financing function with health-system goals – and, in particular, with the fairness-in-financial-contribution goal (Walt and Mills 2001). This criticism was reiterated in the regional consultations and later discussions in which critiques of ideological bias gave way to discussions of how best to improve analysis and measurement of the function in order to generate an objective evidence base for policy advice.

In particular, WHO was requested to develop ways of measuring the effectiveness of different financing mechanisms towards achieving system goals (WHO Regional Office for the Americas 2001; WHO Regional Office for Europe 2001). Through such measurement, it would be possible to 'drill down' and understand why a particular country was performing well or poorly with regard to its health-system goals.

3. WHO responses and proposals

In order to deepen work in this area, WHO has initiated a series of technical papers to review the evidence on well performing financing functions as part of the development of a health financing policy. The technical papers are aimed at a policy audience and seek to synthesize existing evidence on high-priority topics for Member States such as the right amount of funding for health, minimum spending on health, user fees, community financing, and private insurance.

In parallel, WHO has developed a strategy for measuring and characterizing the health financing function with the aim of relating its effectiveness to the health system's intrinsic goals. WHO proposes to undertake pilot activities in a small set of countries in which both "core" variables and country-specific information would be collected and analyzed. This work will be undertaken jointly with national governments, policy makers, and local research institutions. The measures will also be submitted for expert review. As a result of this work, refined "core" variables will gradually be generalized for measurement in other countries.

4. SPRG comments and recommendations

These dimensions and a proposed set of potential variables were discussed with SPRG and are presented in Table 1. They were also discussed at an internal WHO consultation (in March 2002) that included staff from regional and country offices. There was general agreement that the chosen variables should be useful for measuring the function's performance in a variety of dimensions, and should control for important contextual factors. In broad terms, the measurements should help assess how well the system collects, pools, and allocates funds to service provision. This should enable better targeting of appropriate policy actions needed to improve the financing function. SPRG also recommended that WHO draw on the work of other agencies working in this area such as the OECD Health Project (Hurst and Jee-Hughes 2001; Hurst 2002).

In the course of these discussions, the following issues were raised.

- *Minimum threshold of funding for the health sector.* Can WHO provide guidance on what countries should minimally spend on health?
- *Cost of revenue collection.* Any measures of the financing function should include the costs of administering revenue collection because they are a measure of the effectiveness of the collection sub-function.
- *Uncovered population.* The proposed indicators do not appear to measure the number of people who do not belong to any pool, or who are eligible for free public services. This will depend on the definition adopted for "pool", and the kind of data from which the indicators will be derived. It was also noted that people covered by a public "safety net" are implicitly in a "pool". In any case, the indicators should include some measure to reflect this potential failure of a health-financing system.
- *Progressivity of financing.* The financing function is closely tied to the goal of fairness-in-financing. There is concern that the Fairness in Financing Index (FFI) is misleading because the formula can rank an extremely progressive system as being just as 'unfair' as a very regressive one.
- *Differences in benefit packages between pools.* The indicators need to address not only how many people are in the pools, but also what services are covered. When benefit packages differ, the implications for the effectiveness of the financing function will also differ.
- *Risk distribution among pools.* It was noted that the financing function will behave very differently depending on whether there are constraints to adverse selection or exclusion. How will these institutional features be incorporated in the indicators?

- *Overlapping pools.* The indicators will have to take into account the fact that in many countries pools overlap.
- *Payment mechanisms.* The final indicator was unclear. The purpose here will be to analyze how the incentives generated by payment mechanisms affect costs, as well as the amount, kinds and quality of health services provided.
- *Transparency and accountability.* It is important to find ways to measure the transparency and accountability of the financing function since these factors probably have a large impact on its effectiveness.
- *Sources of funds.* Measurements need to capture the wide range of sources for funds flowing into the health sector, including external aid which is significant in many low-income countries.
- *Indicators for research vs. policy guidance.* Questions were raised regarding whether these indicators are sufficiently "universal" to measure cross-country differences. The extent to which they are aimed at supporting research or monitoring policy also requires clarification.
- *Links with other work.* The financing indicators work needs to be well-coordinated with other work on FFC, provision, coverage, and responsiveness.

Recommendations

- WHO should continue to develop measurements of the health-financing function and address the concerns listed above.
- WHO should produce technical papers aimed at consolidating the evidence on the health-financing function and how it affects health-system performance.

5. References

- Almeida, C., P. Braveman, M. R. Gold, C. L. Szwarcwald, J. M. Ribeiro, A. Miglionico, J. S. Millar, S. Porto, N. R. Costa, V. O. Rubio, M. Segall, B. Starfield, C. Travassos, A. Ugá, J. Valente, and F. Viacava (2001): Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357(9269): 1692-1697.
- Häkkinen, U. and E. Ollila, eds. (2000): The World Health Report 2000: What does it tell us about health systems? Analyses by Finnish experts. Helsinki: National Research and Development Center for Welfare and Health (STAKES). [<http://www.stakes.fi/english/publicati/Publications.htm>]
- Helms, R. (2000 August): Sick list: health care à la Karl Marx. *Wall Street Journal*.
- Hurst, J. (2002): Performance measurement and improvement in OECD health systems: Overview of issues and challenges. In *Measuring Up: Improving Health System Performance in OECD Countries*. Paris, France: Organization for Economic Co-operation and Development (OECD).
- Hurst, J. and M. Jee-Hughes (2001): Performance measurement and performance management in OECD health systems. *Labour Market and Social Policy -- Occasional Papers*, No. 47. Paris, France: Organization for Economic Co-operation and Development (OECD).
- Murray, C. J. L. and J. Frenk (2001): World Health Report 2000: A step towards evidence-based health policy. *Lancet*, 357(9269): 1698-1700.
- Navarro, V. (2000): Assessment of the World Health Report 2000. *Lancet*, 356(9241): 1598-1601.
- Navarro, V. (2001a): The new conventional wisdom: an evaluation of the WHO report, Health Systems: Improving Performance. *International Journal of Health Services*, 31(1): 23-33.
- Navarro, V. (2001b): World Health Report 2000: Response to Murray and Frenk. *Lancet*, 357(9269): 1701-1702.
- Oswaldo Cruz Foundation, Brazilian Ministry of Health (2000): Report of the workshop "Health Systems Performance: The World Health Report 2000". 14-12-2000, Rio de Janeiro.
- Van der Stuyft, P. and J. P. Unger (2000): Improving the performance of health systems: The World Health Report as go-between for scientific

evidence and ideological discourse. *Tropical Medicine and International Health*, 5(10): 675-677.

Walt, G. and A. Mills (2001): World Health Report 2000: Comments. *Lancet*, 357(9269): 1702-1703.

WHO Regional Office for the Americas (2001): Regional consultation of the Americas on Health Systems Performance Assessment: Final Report. 8-5-2001, Washington, D.C.

WHO Regional Office for Europe (2001): Report of the Regional Consultation on Health Systems Performance Assessment. 3-9-2001, Copenhagen.

VI. STEWARDSHIP FUNCTION

1. WHR 2000

WHR 2000 introduced the concept of stewardship as one of the four essential functions of the health system: *service provision, resource generation, financing* and *stewardship*. In the Introductory chapter, the Director-General highlighted this new concept noting that the function involves “setting and enforcing the rules of the game and providing strategic direction for all the different actors involved”. The concept was developed further in chapter six.

Here stewardship was defined as “the careful and responsible management of the well-being of the population, the very essence of good government”. The text continued “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”. It also stated that ultimate responsibility for the overall performance of a country’s health system lies with government, which in turn should involve all sectors of society in its stewardship. Within government, Ministries of Health must take on a large part of the stewardship of health systems and should direct/coordinate intersectoral action for health.

2. Main commentaries and criticisms

Although various individuals and groups had commented extensively on aspects of the WHR 2000, there was little published comment on stewardship. Only two articles have been identified. Saltman and Ferroussier-Davis (2000) discuss the concept of stewardship in health policy as proposed in WHR 2000, and conclude that the concept “holds substantial promise if adequately developed and effectively implemented”. An editorial in the *European Journal of Public Health* (McKee 2001a) also discusses the potentially major implications of the concept of ‘stewardship’ of health systems, both for countries and for WHO as it seeks to strengthen its role as a credible advocate for global health.

The Regional Consultations on health-system performance assessment did discuss stewardship. In addition to statements about the importance of the notion of stewardship, it featured in general discussions on the need to better map all the health system functions, their linkages with each other, and their relation to outcomes. Participants in the EMRO consultation observed the need for better definition of each of the components of stewardship (WHO Regional Office for the Eastern Mediterranean 2001). In AFRO and SEARO,

participants recommended that in future work on performance assessment, WHO and Member States should pay special attention to developing methods for assessing the stewardship function of health systems (WHO Regional Office for Africa 2001; WHO Regional Office for South-East Asia 2001). In the EURO consultation, it was stressed that assessment should not be seen as an isolated exercise, but explicitly linked to efforts to strengthen stewardship (WHO Regional Office for Europe 2001). More general comments were also made on the need to develop complementary and qualitative measures of functions, not just pursue quantitative dimensions of analysis. PAHO/AMRO proposed the use of a 'dashboard' approach to assessing functions (WHO Regional Office for the Americas 2001). In the WPRO consultation, the participants commented on WHO's own stewardship role in the international health arena (WHO Regional Office for the Western Pacific 2001).

The most extensive debate on stewardship has been the international technical consultation on stewardship in September 2001 at which the participants reviewed the definition of the term and discussed its relation to governance (World Health Organization 2001). They also noted the difficulty in preserving the idea when translating the term into other languages. Participants referred to it metaphorically as combining three elements ("the 'glue' that holds the health system together; the 'oil' that keeps it running smoothly, and the 'energy' that gives it ethical direction and momentum"). They generated a list of possible stewardship tasks that fitted within the three-part classification that WHR 2000 set out:

- formulating health policy;
- exerting influence; and
- collecting and using intelligence.

Participants agreed that some form of descriptive characterization of approaches to stewardship would be useful, and counselled against measurement of stewardship as an isolated exercise. They emphasized that a clearer understanding is needed of relationships between approaches to stewardship, the resultant effectiveness of the stewardship function, and the performance of health systems.

The few comments on assessment of stewardship have noted that this will be an important yet challenging task. Some have commented on the importance of ensuring that any assessment is useful in the first instance to nationals in their own efforts to improve stewardship.

3. WHO responses and proposals

Stewardship: conceptual issues

Building on the work of other organizations as well as the recommendations of the meeting on the stewardship function held in September 2001, WHO staff prepared a paper in November 2001 that further developed the concept (Travis et al. 2002). The paper tentatively identifies a number of essential ingredients or “core domains” that appear to constitute good stewardship. Domains are conceived as relatively well defined, distinct areas of responsibility that collectively constitute effective stewardship. The six domains or sub-functions that constitute effective health system stewardship, i.e., that lead to better outcomes to achieve the goals of health systems are referred to as:

- Generation of intelligence
- Formulating strategic policy direction
- Ensuring tools for implementation: powers, incentives and sanctions
- Coalition building / Building partnerships
- Ensuring a fit between policy objectives and organizational structure and culture
- Ensuring accountability

The paper states that it is desirable to increase capacity within the health systems of Member States with regard to each of these domains. The assumption is that, collectively, the better these sub-functions are carried out, the more effective health system stewardship will be and the higher attainment of intrinsic goals. These domains, attributes and relationships are based on prevailing notions of effective stewardship, and the paper emphasizes that all should be considered “testable hypotheses”. The concept of stewardship has evolved in the course of the consultations and analyses (see table below):

Three Classifications of Tasks for Stewardship

WHR 2000	Consultation	Travis et al. (2002)
Collecting and using information	Collecting and using intelligence	Generation of intelligence
Defining the vision and direction of health policy	Formulating health policy	Formulating strategic policy direction
Exerting influence through regulation and advocacy	Exerting influence	Ensuring tools for implementation: powers, incentives and sanctions
		Coalition building / Building partnerships
		Ensuring a fit between policy objectives and organizational structure and culture
		Ensuring accountability

The scope and core attributes of each domain are outlined briefly here and described in more detail in the background paper.

Generation of intelligence

This domain responds to the concern "to what extent do health system actors have useful intelligence at their disposal?"; and 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?

Formulating strategic policy direction

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?
- Is there a clear definition of roles and responsibilities of public and private actors in all four functions?
- Has there been a clear identification of policy instruments and institutional arrangements required to achieve improvements?
- Have the authorities developed strategies for making the required changes? and
- Have they provided guidance for prioritizing health expenditures based on realistic resource and needs assessment, and for monitoring effects of changes on performance?

Ensuring tools for implementation: powers, incentives and sanctions

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 or contradictions between stated policy and the regulatory framework (whether powers and responsibilities are matched); and the extent to which they are enforced.

Coalition building/Building partnerships

This domain addresses the concern "to what extent does capacity exist to create alliances of individuals, groups or organizations for joint action around strategic health and health system priorities?"

Ensuring a fit between policy objectives and organizational structure and culture

This domain addresses two questions: "To what extent do organizational structures and management systems fit with policy objectives so that they help rather than hinder policy implementation"; and "to what extent have conditions been created by government that allow stewards themselves to be effective?".

Ensuring 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.

Assessing stewardship

WHO states that the objective of its work is to support health systems' performance improvement at country level by providing evidence-based advice on the relationships between stewardship and health system outcomes. Travis et al. (2002) rightly point out that whilst the importance of many of the activities thought to contribute to effective stewardship have long been written about, as a whole it is a new construct in health systems and there are no tools for looking at all its different aspects. They referred to studies in other areas which have attempted to measure some of the elements of stewardship. In particular, they noted the work done on governance, and work to define an instrument for measuring 'Essential Public Health Functions' (Pan American Health Organization et al. 2001) that may assess selected aspects of stewardship. WHO proposes to explore this and other approaches in the development of credible stewardship assessment tools. 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, including households if appropriate.

Development of an instrument to assess stewardship

Descriptive and analytical approaches will require qualitative and quantitative assessments. The analyses required, audience and intended use will guide the selection of approaches. Therefore, rather than one single instrument, a set of assessment approaches will be developed and tested.

WHO has developed a provisional programme of work to further develop the concept of stewardship and its assessment based on the comments thus far received.

Proposed outputs by December 2003

- Revised conceptual paper and broad-based consensus on domains and assessment indicators.
- Tools for assessing stewardship developed and used in around 15 interested countries.
- Series of country reports on key findings of the stewardship assessment for in-country discussions and policy dialogue.
- Valid and meaningful approach to analyses, for in-country, regional and cross-national use.
- Set of training modules for capacity building on stewardship related to health systems' performance.

- Dissemination and public access to information and linking various research results across functions and goals.

4. SPRG comments and recommendations

Conceptual issues

• **What is the value of this new concept?**

Several commentators (McKee 2001b) see it as an important development. "WHO has clearly stated that governments have a responsibility for their health systems. Stewardship implies a much more active involvement in promoting health than most governments have previously assumed" (McKee 2001a). Not all commentators have seen it as such a departure. In the discussions of SPRG, it was observed that WHO as a technical and political organization has long had an obligation to produce reliable evidence on health or health system issues, and to 'speak up' and publicize that evidence. One WHO Regional Office commentator pointed out that it builds on previous efforts by WHO to strengthen ministries of health and their 'leadership' role. Another noted that there are many examples of current market or political failures arising from the lack of stewardship within ministries of health. Some commentators have requested that WHO provides examples of effective stewardship when different actors in public and private settings have conflicting goals or interests, as a means to illustrate good stewardship more concretely (Saltman and Ferroussier-Davis 2000).

• **The definition of stewardship**

SPRG consider that the concept is well defined. They considered whilst it is not that distinct from governance, the word stewardship may better reflect the element of directing a health system. In the Technical Consultation, the ethical foundation of the concept was stressed (World Health Organization 2001). One commentator observed that it is the 'moral' aspect of stewardship that distinguishes it most from governance, which is seen as more procedural notion. Another distinction has been drawn between stewardship as an 'intelligent' function and governance as a more structural one – a set of activities that have to happen.

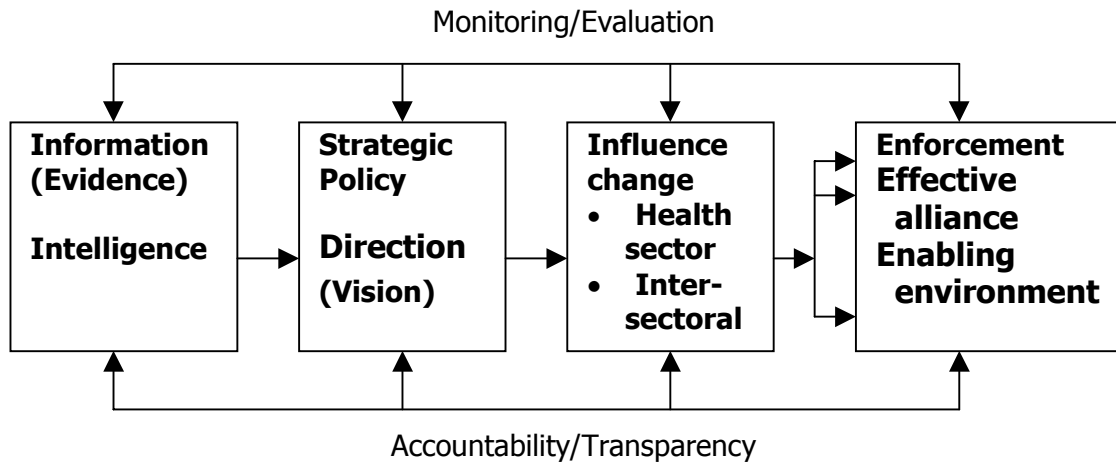
• **Who is responsible for stewardship?**

Governments have the primary responsibility for discharging the stewardship function of the health system. The expectation is that each governments will ensure that it meets the legitimate health needs and expectations of the population in the context of available resources. This does not imply that the government will be solely responsible for performing all the essential tasks. In discharging its stewardship role, the government develops partnerships,

works with civil society and with the private sector but such linkages cannot dilute the fact that governments are primarily responsible for discharging the function of stewardship. Active participation by the civil society can be of great value in developing national goals and in ensuring good stewardship by government. The active involvement of the civil society and the contributions of the private sector are vital components of the health system but the people have the right to hold their governments accountable for the operation of the health system as a whole. Responsibility refers who must ensure that justifiable expectations are met and it identifies who must take the blame when things go wrong.

The proposed domains/sub-functions of stewardship

SPRG supports the six elements of stewardship and proposes an interconnected framework for these six elements:



More work is required to characterize each domain more clearly. Aspects of stewardship that are currently insufficiently addressed or unclear include:

- The need for a clearer link between the generation and use of intelligence.
- The need to be clear that stewardship is not only about central control.
- The early warning / detection function of stewardship – for example of harmful practices.
- The refereeing function - detection and dealing with conflicts of interest.

Assessment issues

In the SPRG discussion, WHO emphasized that the work being proposed was mainly aimed at improving understanding of the different components of stewardship, so that action could be taken by countries. Qualitative as well as quantitative approaches are likely to be needed.

On the question about the use of surveys to measure stewardship, SPRG raised several issues:

- Who should be surveyed – key informants, households, or a mix?
- Who should carry out such surveys?
- Would assessment from household surveys be useful to governments who may have to make unpopular decisions?

The reliability, validity and comparability of any survey approach were viewed as key issues in ensuring acceptability and use of results. A note of caution was provided by SPRG that whatever aspects of stewardship are selected for assessment, these will automatically be assumed to be important or desirable, when there is still little evidence.

SPRG commends the Secretariat's initiative to measure the level of stewardship. The Group suggests that in addition to subjective measurements, some objective indicators should also be developed. The key stakeholders who will be the respondents should be clearly defined and identified and conflict of interests avoided. A composite index on stewardship may be developed. However, as this is quite a sensitive area involving the function and responsibility of national health leaders, great care should be accorded to the methodology, the measurements, and the publication.

5. References

McKee, M. (2001a): The challenge of stewardship. *European Journal of Public Health*, 11: 122-123.

McKee, M. (2001b): Measuring the efficiency of health systems. *British Medical Journal*, 323(7308): 295-296.

Pan American Health Organization, World Health Organization, Centers for Disease Control and Prevention, and Centro Latinoamericano de Investigaciones en Sistemas de Salud (2001): "Public Health in the Americas" Initiative: Instrument for Performance Measurement of Essential Public Health Functions.

Saltman, R. B. and O. Ferroussier-Davis (2000): The concept of stewardship in health policy. *Bulletin of the World Health Organization*, 78(6): 732-739.

Travis, P., D. Egger, P. Davies, and A. Mechbal (2002): Towards better stewardship: concepts and critical issues. Global Programme on Evidence for Health Policy Discussion Papers, No. 48. Geneva, Switzerland: World Health Organization.

World Health Organization (2001): Report on WHO meeting of experts on the stewardship function in health systems. 10-9-2001, Geneva, Switzerland. (unpublished document HFS/FAR/STW/00.1).

WHO Regional Office for the Eastern Mediterranean (2001): Report on the Regional Consultation on the Conceptual Framework for Health System Performance Assessment. 9-7-2001, Ain Saadeh, Lebanon.

WHO Regional Office for Africa (2001): General Report on the Regional Consultative Meeting on Health Systems Performance Assessment: Final Report. 18-5-2001 Harare, Zimbabwe.

WHO Regional Office for the Americas (2001): Regional consultation of the Americas on Health Systems Performance Assessment: Final Report. 8-5-2001, Washington, D.C.

WHO Regional Office for Europe (2001): Report of the Regional Consultation on Health Systems Performance Assessment. 3-9-2001, Copenhagen.

WHO Regional Office for the Western Pacific (2001): Report of the regional consultation on health system performance. 3-7-2001, Manila, Philippines.

REPORT OF SPRG ON HSPA

WHO Regional Office for South-East Asia (2001): Report of the regional consultation and technical workshop on health systems performance assessment. 18-6-2001, New Delhi, India.

VII. AVERAGE LEVEL OF POPULATION HEALTH

1. WHR 2000

WHO has reported indicators of population health for each Member State for many years, including child and adult mortality risks and life expectancy at birth. In addition, for over a decade WHO has been involved in the development of summary measures of population health (SMPH), which combine information on mortality and non-fatal health outcomes to represent population health as a single number. WHR 2000 reported disability-adjusted life expectancy (DALE) at birth and at age 60, for males and females, with uncertainty intervals around the most likely estimates.

As part of this process, new life tables and life expectancies were estimated for all 191 Member States. Data were taken from different sources. For example, vital registration data were available for 80 countries. In other countries, indirect information on mortality – infant, child and/or adult – was available. At times, UN Population Division estimates of adult mortality were used where no direct sources of data were available. For countries without adequate vital registration data or surveys, estimates were based on regional logit models.

To estimate DALE for each country, the life table data were supplemented with information on age- and sex-specific prevalences of non-fatal health outcomes and 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 (GBD) estimates. Existing health surveys were also used, but the additional information they provided was limited by problems of cross-population comparability.

2. Main commentaries and criticisms

Summary Measures of Population Health

Much of the discussion about the indicator of the level of health used in WHR 2000 was a continuation of the long-standing debate about the value of summary measures of population health (SMPH). For example, it was argued that SMPH do not describe health in sufficient detail to be useful for policy makers. Reporting the components separately is of more value, e.g. mortality, and prevalence, incidence duration and severity of various non-

fatal health outcomes (Navarro 2001a; Rosén 2001; WHO Regional Office for Africa 2001). Some commentaries criticized disability-adjusted life years (DALYs), stating that by themselves they should not be used for resource allocation decisions (Almeida et al. 2001; Häkkinen 2000; Rissanen and Sintonen 2000). SMPH were seen to be too complex for policy makers to understand (Almeida et al. 2001; Oswaldo Cruz Foundation 2000) and the fact that Life Expectancy at birth (LE) and Disability-Adjusted Life Expectancy (DALE) were highly correlated led some critics to argue that DALE added little that was not already captured by LE (Oswaldo Cruz Foundation 2000; Häkkinen 2000, McKee 2001a; Ugá et al. 2001).

Another concern with SMPH surrounded the valuation of non-fatal health outcomes on the same scale as death and full functioning. Critics of summary measures argued that this type of assessment undervalued the lives of disabled people (Almeida et al. 2001; Oswaldo Cruz Foundation 2000) and also raised other ethical problems (Oswaldo Cruz Foundation 2000; Rissanen and Sintonen 2000; Nord 2002). Others claimed that the valuations used in WHR 2000 did not capture all aspects of quality-of-life or heterogeneity across countries in the way people understood and valued health (DfID 2000; Oswaldo Cruz Foundation 2000; Rissanen and Sintonen 2000; WHO Regional Office for the Americas 2001; Nord 2002). Still others argued that weights should ideally be obtained from representative population groups rather than from a limited group of experts (Almeida et al. 2001; Oswaldo Cruz Foundation 2000; Rissanen and Sintonen 2000; WHO Regional Office for the Eastern Mediterranean 2001).

The word "disability" in DALE also raised concerns: "disability" could be seen as a pejorative term and was not an appropriate word to use to describe a state that is less than full health. Moreover, it does not well capture the idea that health is a multidimensional and complex concept (Oswaldo Cruz Foundation 2000; Häkkinen and Ollila 2000; Van der Stuyft and Unger 2000).

As a key goal in assessing the performance of health systems, a number of authors pointed out that measures such as DALE reflected past as well as current performance, and hence cannot be interpreted as being a function only of current performance (Almeida et al. 2001; DfID 2000; McKee 2001b; Oswaldo Cruz Foundation 2000; Rosén 2001; Häkkinen 2000; WHO Regional Office for Europe 2001; Ministry of Health, Vietnam 2001; McKee 2001a).

A series of specific technical points were also raised about the construction of DALE. For example, life tables estimated for countries where vital registration data did not exist do not fit the oldest age groups well in some countries (WHO Regional Office for Africa 2001; WHO Regional Office for the Americas 2001; WHO Regional Office for the Eastern Mediterranean 2001; WHO Regional Office for Europe 2001; WHO Regional Office for South-East Asia 2001; WHO Regional Office for the Western Pacific 2001). The way in which the estimates of the prevalence of non-fatal health outcomes were obtained, and co-morbidity was handled in developing the overall severity-

adjusted prevalence of non-fatal health outcomes, was seen as simplistic (Navarro 2001a; Rosén 2001; Rissanen and Sintonen 2000; World Health Organization 2001). For example, the prevalence of different types of disability was assumed to be the same in all countries with similar life expectancies (Oswaldo Cruz Foundation 2000; Ugá et al. 2001). In addition, McKee (2001a) argued that in estimating uncertainty intervals around the estimates, all possible sources of uncertainty had not been considered.

The final set of commentaries concerned data sources. It was generally agreed that epidemiological data were sparse and of variable quality in many countries, and that the use of more vital registration data would greatly improve the estimates (McKee 2001b; Oswaldo Cruz Foundation 2000; Williams 2001; Häkkinen 2000; Rissanen and Sintonen 2000; WHO Regional Office for Africa 2001; WHO Regional Office for the Eastern Mediterranean 2001; WHO Regional Office for Europe 2001; Ugá et al. 2001). The available data on child and adult mortality, as opposed to infant mortality, were particularly poor (WHO Regional Office for Africa 2001). WHO had used UN Population Division estimates at times, and the sources and methods underlying them were seen to be unclear (WHO Regional Office for Europe 2001; McKee 2001a).

3. WHO responses and proposals

Noting that there are large variations in DALE for any given level of LE, the WHO Secretariat argued that DALE does indeed provide additional information to that contained in LE. It proposed to SPRG that it should continue to use SMPH to measure and monitor population health. In doing this, it was important to be clear that the question of measuring and monitoring population health was not the same as the question of resource allocation. The criticism that DALYs should not be used for resource allocation was not relevant to this debate – as this was not being proposed. To help policy makers identify the possible causes of changes in health outcomes, WHO proposes to continue publishing the components of DALE separately – i.e., mortality and non-fatal health outcomes. SMPH is a complement to, but not a substitute for, information on the separate components.

In recognition of the problems associated with the term “disability”, and the fact that health is a multidimensional concept, WHO proposes to accept the advice that the name of the indicator should be changed from DALE to health-adjusted life expectancy (HALE).

To respond to criticisms related to health valuation a more precise definition of its conceptual basis was provided. Part of the WHO Multi-Country Survey Study 2000-2001 involved detailed questionnaires in 12 countries designed to

explore if people from different cultures rated the domains of health differently. WHO used these results to develop a method for estimating new health-state valuations (Salomon et al. 2002). The Secretariat proposes to apply it to the data from the World Health Survey (WHS): a global average valuation function will be applied to the individual domain levels estimated using the HOPIT model (Section XIII) in order to derive severity-adjusted prevalences of health states by age and sex for each survey country. At present, the prior estimates of severity-weighted prevalences derived from the GBD study use the disability weights from GBD 1990, together with weights from the Dutch disability weights study (Stouthard et al. 1997). WHO plans to revise all disability weights used in the GBD study from the population-based valuations that are obtained from WHS.

The question of timing is taken up again in Section XIV on 'Efficiency' in the SPRG report. To complement the prevalence-HALE that is already reported routinely, WHO proposes to explore the feasibility of developing an incidence-HALE based on current incidence and transition rates and information on current exposures to major risk factors. It would then be determined largely by actions undertaken in the current time period.

WHO has undertaken intense efforts to obtain more and better data. This started with consultations between WHO and Member States to verify the best sources of recent data on vital registration and causes of death. The number of countries with relatively complete vital registration or cause-of-death data has increased from 80 in 2000 to 110 in 2002. New life tables for the year 2000 have been constructed for all 191 Member States using these data. A modified logit life table model was developed for countries with incomplete registration or survey data. It employed a much larger empirical database of observed life tables than any previous model life table system, and has resulted in much better estimates of mortality at older ages. Separate estimates of HIV mortality were made for countries with high HIV mortality.

In addition to methods development for life tables, WHO has introduced ways of taking co-morbidity into account in estimating HALE. This was facilitated partly by the Multi-Country Survey Study, which included instruments and analytical tools for improving cross-population comparability of survey data. Comparable data on the prevalence of non-fatal conditions from 63 surveys in 55 countries were used to estimate HALE for WHR 2001. (The new statistical methods to establish cross-population comparability are discussed in Section XIII.) The other components of the calculations were:

- (a) Direct estimates of prevalence for major disease and injury sequelae.
- (b) Country-level prevalence data for selected conditions.
- (c) Regional information, specific epidemiological studies, and available country information on cause-specific mortality to estimate morbidity in countries with poor information about causes.
- (d) Adjustment by known under-registration for highly stigmatized causes of morbidity and mortality such as abortion, HIV/AIDS, and suicide.

- (e) For estimating health-state prevalences, data from the Multi-Country Survey Study were used (cross-population comparable prevalences, and valuations based on population preferences) together with severity-weighted prevalences derived from epidemiological analyses in GBD 2000. In addition, improved Bayesian methods were used to compute posterior health-state prevalences that combine GBD 2000-based 'prior' estimates with prevalence estimates from the Multi-Country Survey Study. For those countries with no survey results, a relationship between posterior and prior estimates of prevalences for the survey countries was used to update the priors.

Finally, improved methods for uncertainty analysis were used, including more explicit and comprehensive treatment of uncertainty in various inputs. The uncertainty interval of 80% used in WHR 2000 was increased to 95% in WHR 2001. WHO proposes to continue calculating and reporting uncertainty intervals in a systematic manner so that different users can make their own assessment of the estimates.

4. SPRG comments and recommendations

- (i) WHO is playing a leading role in the development of new concepts and health measures that incorporate non-fatal outcomes into SMPH. Considering the complexity of the issues, it is natural that there will be debate about these innovations – and some policy makers will prefer to use indicators of individual components of health rather than summary measures. Moreover, SMPH are sometimes seen as having less validity than the single component measures, especially for Member States where both morbidity and mortality estimates have wide uncertainty intervals. WHO should continue to emphasize that SMPH complement rather than compete with the disaggregations of the component parts, and it should continue to take steps to make more detailed disaggregations of SMPH available.
- (ii) SMPH require valuation of health outcomes to allow non-fatal conditions to be combined with mortality. WHO should take additional steps to explain and clarify the concept of health-adjusted life expectancy. This should be distinguished from the complexity of the methods needed for estimation where appropriate data are not available, and from the issue of cross-population comparability.
- (iii) Despite these difficulties, WHO should continue to improve the conceptual and technical aspects of health measurement, engage in external debate and consultation, and obtain better data.

- (iv) SPRG recommends that WHO take steps to strengthen local capacity to build and use these measures, particularly in developing countries. Related to this, SPRG believes that it would be valuable for WHO to establish a permanent forum for discussion of conceptual and methodological aspects of health measurement, and promote the participation of academics, policy makers and civil society – especially from developing countries. As part of this process, there should be a continuing dialogue with social scientists from subject areas such as ethics, anthropology and sociology, so as to take into account insights from these disciplines on the 'value' of health.
- (v) The modified logit life tables provide a reasonable methodology for countries where vital registration data are not available, but the assumptions behind their construction and use should be made more comprehensible for non-expert audiences.
- (vi) Vital statistics registration (VSR) systems are complex and expensive, and do not exist in many developing countries. WHO should encourage the establishment of these registries and provide the necessary technical assistance. Considering the inevitable time lags to establish a functioning VSR system, the use of indirect methods is acceptable as an intermediate solution.
- (vii) In relation to WHS, SPRG believes it is important to increase the number of participating countries, especially those with inadequate health-information systems. The face validity of the WHS data is still an issue. Although the inclusion of vignettes in the questionnaire facilitates the comparability of self-report data between countries, further development and testing of the methods is recommended.
- (viii) Data on adult mortality are still scarce. The WHO Multi-Country Survey Study 2000-2001 found that questions on deaths in households provided some useful information, but that there was underreporting of deaths. An expanded module on adult mortality should be included in WHS and validated in countries with good vital registration data. Improved methods should be developed to maximize the usefulness of this information for estimation of adult mortality.
- (ix) It is important that WHO clarifies the methods and procedures used to estimate causes of death by age and sex for countries without vital registration data or with only partial data, and that it intensifies data collection efforts in such countries.
- (x) HALE incorporates prevalences and valuations of health states from population surveys and from GBD 2000. Currently, the weights used in the GBD 2000 study are predominantly based on the GBD 1990 weights with some additional weights from a Dutch study (Stouthard et al. 1997). Examination of GBD 1990 and Dutch weights (Mathers et al.

1999) suggest that these weights are comparable. SPRG welcomes the effort made by WHO to improve the health-state valuation methods and endorses the proposal to revise the GBD disability weights using valuations derived from the forthcoming WHS.

- (xi) The difference between the uncertainty analysis proposed by WHO and statistical confidence intervals should be made clear in WHO publications.
- (xii) National simulations based on regional or global estimates can be a good starting point to encourage National Burden of Disease studies.

5. References

Almeida, C., P. Braveman, M. R. Gold, C. L. Szwarcwald, J. M. Ribeiro, A. Miglionico, J. S. Millar, S. Porto, N. R. Costa, V. O. Rubio, M. Segall, B. Starfield, C. Travassos, A. Uga, J. Valente, and F. Viacava (2001): Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357(9269): 1692-1697.

DfID (Department for International Development, UK Health Systems Resource Center) (2000): World Health Report 2000: Summary and Comments.

Häkkinen, U. (2000): Assessment of the goal attainment and efficiency of health systems in the World Health Report 2000. In eds. U. Häkkinen and E. Ollila. *The World Health Report 2000: What does it tell us about health systems? Analyses by Finnish Experts*. Helsinki: National Research and Development Center for Welfare and Health (STAKES). [<http://www.stakes.fi/english/publicati/Publications.htm>]

Häkkinen U. and E. Ollila, eds. (2000): *The World Health Report 2000: What does it tell us about health systems? Analyses by Finnish experts*. Helsinki: National Research and Development Center for Welfare and Health (STAKES). [<http://www.stakes.fi/english/publicati/Publications.htm>]

Mathers, C., T. Vos, and C. Stevenson (1999): *The burden of disease and injury in Australia*. Canberra, Australia: Australian Institute of Health and Welfare.

McKee, M. (2001a): The World Health Report 2000: Advancing the debate. Prepared for the European Regional Consultation on Health Systems Performance Assessment. 3-9-2001, Copenhagen.

McKee, M. (2001b): Measuring the efficiency of health systems. *British Medical Journal*, 323(7308): 295-296.

Ministry of Health, Vietnam (2001): Comments and suggestions of Vietnam Ministry of Health/Health Policy Unit as regards the World Health Report 2000.

Navarro, V. (2000): Assessment of the World Health Report 2000. *Lancet*, 356(9241): 1598-1601.

Navarro, V. (2001a): The new conventional wisdom: an evaluation of the WHO report, Health Systems: Improving Performance. *International Journal of Health Services*, 31(1): 23-33.

Navarro, V. (2001b): World Health Report 2000: responses to Murray and Frenk. *Lancet*, 357(9269): 1701-1702.

Nord, E. (2002): Measures of goal attainment and performance: A brief, critical consumer guide. *Health Policy*, 59(3): 183-191.

Oswaldo Cruz Foundation, Brazilian Ministry of Health (2000): Report of the workshop "Health Systems Performance: The World Health Report 2000". 14-12-2000, Rio de Janeiro.

Rissanen, P. and H. Sintonen (2000): Measurement of the state of health in the World Health Report 2000. In eds. U. Häkkinen and E. Ollila. *The World Health Report 2000: What does it tell us about health systems? Analyses by Finnish Experts*. Helsinki: National Research and Development Center for Welfare and Health (STAKES).

[<http://www.stakes.fi/english/publicati/Publications.htm>]

Rosén, M. (2001): Can the WHO Health Report improve the performance of health systems?. *Scandinavian Journal of Public Health*, 29(1): 76-80.

Salomon, J., C. J. L. Murray, B. Ustun, and S. Chatterji (2002): Health state valuations in summary measures of population health. Global Programme on Evidence for Health Policy, mimeo. Geneva, Switzerland: World Health Organization.

Stouthard, M. E. A., M. L. Essink-Bot, G. J. Bonsel, J. J. Barendregt, P. G. N. Kramer, van de Water, H. P. A., Gunning-Scheppers L. J., and P. J. Van der Maas (1997): *Disability Weights for Diseases in the Netherlands*. Rotterdam, The Netherlands: Department of Public Health, Erasmus University.

Ugá A. D., C. M. Almeida, C. L. Szwarcwald, C. Travassos, F. Viacava, J. M. Ribeiro, N. R. Costa, P. M. Buss, and S. Porto (2001): Considerations on methodology used in the World Health Organization 2000 Report. *Cadernos de Saude Publica*, 17(3): 705-712.

Van der Stuyft, P. and J. P. Unger (2000): Improving the performance of health systems: the World Health Report as go-between for scientific evidence and ideological discourse. *Tropical Medicine and International Health*, 5(10): 675-677.

Williams, A. (2000): Science or marketing at WHO? A Commentary on 'World Health 2000'. *Health Economics*, 10(2): 93-100.

Williams, A. (2001): Science or marketing at WHO? Rejoinder from Alan Williams. *Health Economics*, 10(4): 283-285.

REPORT OF SPRG ON HSPA

World Health Organization (2001): Report on WHO meeting of experts on statistical methods for enhancing the cross-population comparability of survey results. 1-10-2001, Cambridge, MA, USA.

WHO Regional Office for Africa (2001): General Report on the Regional Consultative Meeting on Health Systems Performance Assessment: Final Report. 18-5-2001, Harare, Zimbabwe.

WHO Regional Office for the Americas (2001): Regional consultation of the Americas on Health Systems Performance Assessment: Final Report. 8-5-2001, Washington, D.C.

WHO Regional Office for the Eastern Mediterranean (2001): Report on the Regional Consultation on the Conceptual Framework for Health System Performance Assessment. 9-7-2001, Ain Saadeh, Lebanon.

WHO Regional Office for Europe (2001): Report of the Regional Consultation on Health Systems Performance Assessment. 3-9-2001, Copenhagen.

WHO Regional Office for South-East Asia (2001): Report of the regional consultation and technical workshop on health systems performance assessment. 18-6-2001, New Delhi, India.

WHO Regional Office for the Western Pacific (2001): Report of the regional consultation on health system performance. 3-7-2001, Manila, Philippines.

VIII. HEALTH INEQUALITY

1. WHR 2000

WHR 2000 defined 'total health inequality' as 'inter-individual variation in healthy life expectancy', thus basing inequality assessment on between-individual and not between-group variation in health expectancy. An important conceptual characteristic of this approach arises from the fact that an individual's health expectancy cannot be observed, but must be estimated.

The Report argues that the ideal approach is to combine individual risks of ill-health and death across ages in a measure of healthy life expectancy, and summarize the distribution of these risks into a measure of inter-individual inequality. However, owing to lack of international data on individual risks across the age groups, WHR 2000 was only able to estimate inequality in the probability (duration) of child survival to age 2.

The index of inequality used in WHR 2000 is as follows:

$$II[\alpha, \beta] = \frac{\sum_{i=1}^n \sum_{j=1}^n |h_i - h_j|^\alpha}{2n^2 \bar{h}^\beta}$$

where h_i is the expected survival time for child i , n is the number of children in the population, and \bar{h} is the average expected survival time for all children. The alpha parameter is derived from interviews aimed at assessing aversion to inequality, and the beta parameter is similarly derived from individual preferences for a relative versus absolute measure of inequality. The alpha and beta parameters were estimated from information obtained through internet interviews of approximately 1,600 persons.

The above index of inequality was applied to child survival to age 2, and is called the *Index of Child Survival Inequality*. The individual survival rates and risk profiles for children were estimated from maternal birth histories and other covariates, using the Demographic and Health Surveys (DHS) database for 50 countries. As mentioned earlier, no measures of adult health inequality were developed for WHR 2000.

2. Main commentaries and criticisms

Total or partial health variation

A concern with the concept of total health inequality is that it includes all variation in health in a population, without making any judgements as to which part of the variation is unfair. For example, during the technical consultation on health inequalities it was debated whether voluntary or genetic risks should be excluded from the assessment of total variation, indicating a discomfort with the notion that all inter-individual variation is unfair.

Inter-individual and/or social group approaches to inequality

The inter-individual approach to inequality in WHR 2000 has generated impassioned debate about the appropriateness and relevance of inter-individual versus social-group inequality measurement. A number of analyses (Braveman et al. 2001; Houweling et al. 2001; Ugá et al. 2001; Szwarcwald 2002) have shown the relative independence of the social-group measures of inequality from the index reported in WHR 2000, and have argued for both social group and inter-individual assessments of inequality.

Inequality in risks of healthy life expectancy

It is not clear whether the methods employed by WHO to estimate the underlying distributions of risk are applicable to settings where there are no data at the individual level. Wolfson and Rowe (2001) raise concerns about using small geographic area data in these models, as the models are based on the assumption that the population within the areas is homogeneous; in cases where this assumption does not hold, these methods should not be used.

Index of child survival inequality

Szwarcwald (2002) emphasizes the fact that for a majority of countries the index of child survival inequality was not based on child survival data, but estimated using a regression model. This is viewed as a major weakness of WHR 2000.

The choice of the age group (survival up to age 2) has been justified on the basis that mother's recall of child survival beyond age 2 may be defective, and that the survival risks and their distribution are not significantly different for older children (up to age 5). However, there are significant differences in the causes and risks of child death at different ages (neo-natal, infant, and child) and at different levels of overall mortality (high and low), thus calling into question the appropriateness of this age-grouping to capture inequality in child survival.

As the index of child survival inequality (a generalized Gini coefficient) is homogeneous of degree 2.5, it matters whether survival time is measured in months or years (Szwarcwald 2002). Moreover, defining the index of child survival equality as the complement of the index of child survival inequality (1 minus the index of inequality) would be strictly correct only if the latter is a relative measure (Szwarcwald 2002). Both these considerations affect the scale used to measure inequality.

As the index is a generalized Gini coefficient, it will not be decomposable into additively separable components (with the between- and within-components adding up to total inequality). This makes the measured contributions of the different 'components' of inequality dependent on the order in which they are introduced into the decomposition exercise (e.g., holding income constant first and then education, or vice-versa). Hence the magnitude of the components will be difficult to assess unambiguously.

The empirical values of the inequality measure demonstrate a very tight range across countries with low child mortality. However, it has been remarked in the literature that significant residual social-group inequalities in child survival do exist in these low mortality countries but are not captured by the WHO inequality index (Houweling et al. 2001; Leon et al. 2001; Szwarcwald 2002).

The specification of the risks of child survival that are related to maternal characteristics may not include important residential and environmental covariates. As noted by Wolfson and Rowe (2002), the cross-sectional DHS data may not be of sufficiently high quality to estimate inequality in risks.

DHS data are used to derive the measure of child survival inequality in the year 2000. However, DHS data are collected infrequently in most countries, and the year 2000 estimates of inequality cover a wide range of years (e.g., 1975-1985, 1985-1995). Szwarcwald (2002) raises questions about what these estimates actually represent in terms of the data under consideration, and about what the realistic time interval should be for the calculation of inter-temporal change.

As pointed out by Braveman et al. (2001), Houweling et al. (2001), Leon et al. (2001), Ugá et al. (2001) and Szwarcwald (2002), the policy value of the WHO measure of 'total health inequality' relies on an analysis of its determinants, which was not included in WHR 2000.

Key technical issues

Equality standard

The current WHO approach to health inequality measurement is based on the total inter-individual variation in health for a population. The unresolved

question is whether there is an appropriate common fairness standard for all countries against which to assess this variation.

Alpha parameter – aversion to inequality

The values of this parameter are likely to be dependent on the age group and type of health outcome under consideration. For example, populations may have more aversion to a certain level of inequality in children's health compared with a similar level of inequality in adults' health. Similarly, the alpha parameter may be more sensitive to inequality in survival than in states of (ill-) health.

Beta parameter – absolute versus relative measures

It is important to note that the WHO measure of total health inequality falls between a purely absolute and a purely relative measure. As such it is different from most measures used in the literature and WHO should be clear as to its interpretation.

Estimation of alpha and beta

On the empirical side, a more transparent estimation procedure for the alpha and beta parameters is desirable. At present, it is not clear that the information obtained through the internet surveys uniquely identifies a person's alpha and beta parameters in the inequality formula.

3. WHO responses and proposals

Total or partial health variation

WHO suggests that social-group inequalities in health outcomes ignore the within-group inequalities that exist in social groups, and argues that the poor health expectancy of individuals should be of concern independently of their membership of a social group. Therefore, WHO proposes to continue to measure overall or total health inequality in a population, ideally in terms of inequality in health-adjusted life expectancy (HALE). In addition, it proposes to introduce a special focus on the health of the poor, using data on assets ('permanent income') from the World Health Survey to identify the poor.

WHO has intensified efforts to increase the availability of data through multiple sources including DHS, the Pan-Arab Project on Children, the UNICEF Multiple Indicator Cluster Surveys (MICS), and an abridged birth history module in WHS (for countries that do not have a recent DHS). Relevant birth history data from up to 120 countries should be available for the next round of estimation of inequality in child survival.

The covariates of risk of child mortality have been used in a 'decomposition' analysis to identify the effect of changing one covariate on the inequality

index, e.g. removing inequality in income or improving education levels. This work represents an explicit response by WHO to the critique concerning the policy relevance of measuring pure inter-individual inequality. The analysis attempts to identify the main sources of inequality in a population with a view to suggesting policies and interventions to reduce these inequalities. For example, improving access to health care appears to lower the inequality index of child survival.

Inequality in adult health

In response to the criticisms about the lack of relevance of the indicator for high-income countries, WHO is proceeding with the development of methods to estimate inequality in healthy life expectancy. A survival analysis model has been developed to estimate the distribution of adult mortality risk. This model is similar to the one used for children, and can be used on individual level data from health surveys and censuses that have been linked to vital registration information in some Member States. The survival analysis model includes a shared frailty component that is able to capture unmeasured community effects on adult mortality.

This approach, while appropriate for high-income countries with sophisticated vital registration systems and computerized census and health survey databases, is difficult to apply in the majority of Member States. For Member States where individual-level data are not available, statistical techniques are being developed by WHO to estimate the distribution of mortality risk from small-area data. Wolfson and Rowe (2002) and Szwarcwald (2002) note that the approach of using small-area data depends on the construction of the geographical areas and the homogeneity of the population in them. Inevitably, small-area data will underestimate the true level of inequality. In order to quantify the extent of under-estimation, WHO proposes to compare results from small-area data analysis with results from individual-level data analysis for about 10 Member States where both types of data are available.

Statistical models and micro-simulation techniques developed and already in use at Statistics Canada will help in the implementation of this strategy. If a systematic relationship is found between the estimates from the two types of data, small-area data can be used with more confidence to estimate the distribution of mortality risk for adults where individual-level data are not available. In this case, WHO would proceed with small-area analysis of health inequality in a number of Member States (approximately 50 to 60) where vital registration and health data exist for relatively homogeneous small areas, such as municipalities or counties.

Inequality in health states

In WHR 2000 there were no measures of health inequality related to non-fatal health outcomes. This not only reflects the absence of data for such an

assessment but also the challenge of identifying an appropriate indicator of non-fatal health that is amenable to inequality measurement.

In its Multi-Country Survey Study 2000-2001, WHO has collected data on health states for nationally representative samples of males and females of all ages in a manner that allows for cross-country comparisons. WHO proposes to incorporate the distribution of health states by age and sex into the estimation of HALE for Member States. Preliminary results from the analysis of these data suggest that the observed trends in survival, where there is more variation for males than for females, are different from trends in non-fatal health outcomes, where there is more variation for (adult) females than for (adult) males. This highlights the importance for WHO to complete this analysis and continue with the implementation of WHS for reporting on health inequality in the future.

4. SPRG comments and recommendations

Total or partial health variation

The concept of total health inequality defined as 'inter-individual variation' raises the question of the purpose of measuring inequality in health. Are health systems interested in assessing distributional performance by describing total variation (which is perhaps an over-inclusive notion of inequality)? Or are health systems interested in assessing distributional performance by describing inequalities that are thought to be unfair (which are a sub-set of the total variation)? SPRG recommends that WHO should continue to foster open debate on these two approaches.

Single or multiple measures of inequality

We acknowledge that WHO needs to use a single measure of inequality in its final estimation of health-system attainment. But as the average level of population health is reported separately by WHO, this measure can be a purely relative one. In addition to the single measure used in HSPA, WHO should also report on health inequality using alternative summary measures such as the inter-quartile range, the Gini coefficient, and the coefficient of variation. A combination of measures may need to be calculated to encompass concerns about distinct aspects of inequality.

Social group measures of inequality

SPRG endorses WHO plans to estimate separately the health of the poor and the non-poor in Member States. SPRG also recommends that a broader range of social inequalities in health be assessed, including gender and racial (or ethnic) inequalities. These social groupings raise fundamental issues

related to the norms against which the inequalities are assessed, e.g. genetic differentials in survival between the sexes. WHO should take account of the current policy environment to assess which of these group stratifiers is most useful in identifying inequalities. For example, the pervasive move towards decentralized health systems raises the importance of being able to identify both within- and between-district inequalities in health. SPRG noted that there was no conceptual or empirical attempt to assess gender inequalities in health in WHR 2000.

Inequality in individual risks

Further validation of the approach proposed by WHO to measure inequality in the distribution of health expectancy is recommended. It should be sensitive to the extreme ranges in levels of mortality across countries, and not overly dependent on the level of healthy life expectancy.

Index of child survival inequality

A number of technical recommendations on the index of child survival inequality follow:

- More explicit deliberation is needed about the equality standard against which inequality in health performance is assessed.
- Further estimates of the alpha parameter used to incorporate aversion to inequality are required for different age groups and for different health outcomes.
- Instead of relying exclusively on a hybrid absolute-relative measure, separate indices of relative and absolute inequality would help to clarify the different impact they have on assessing both the level of, and trends in, inequality.
- The robustness of the current index should be evaluated further by comparing it to more tightly defined age-group measures of inequality, e.g. neo-natal, infant, and child.
- The sensitivity of the index to inequality in high-, medium- or low-mortality settings should be explored, and if necessary an index developed that is appropriately sensitive.
- The current risk model (based on cross-sectional DHS data) should be validated by using longitudinal data on child survival from demographic surveillance sites such as Matlab, Bangladesh or Navrongo, Ghana.
- Approaches to reporting the inequality index for estimates based on data collected in different years and in different countries, but reported for a single year (e.g., 2000), should be standardized.
- Given the rate at which new child survival data are generated, guidelines for the frequency of reporting on the 'Index of Inequality in Child Survival' should be developed.

Decomposition of the inequality index

Further work should specify whether the models of the determinants of inequality in child survival perform differently from the models of the determinants of average child survival. WHO should take advantage of longitudinal data to develop more robust models of the determinants of inequality to understand better the extent to which modifying risk factors alter inequality.

Adult health inequality

SPRG recommends that WHO proceeds with the evaluation of small area variation as a basis for deriving reliable estimates of adult inequality.

Data requirements

Given the number of countries in WHR 2000 for which data were missing for the estimates of child survival inequality, it is likely that an absence of data will limit more ambitious inequality measurement efforts, e.g. for a broader range of age-sex groups and health states. The data intensity of these methods raises significant concerns about whether countries, especially those with limited health-information systems, will be able to invest in such data. WHO should propose strategies for the sustainability of the assessment of health inequality at regular time intervals.

General recommendations for future development

Although fairness in the distribution of health is a key performance criterion of health systems, it is not clear that the current 'Index of Inequality in Child Survival' actually informs the fairness in health outcomes component of HSPA. The literature, experts and members of SPRG have raised a considerable number of conceptual and technical challenges, some of which have been addressed, while others are only beginning to be explored. For these reasons SPRG believes that the health inequality aspect of HSPA would benefit from further conceptual, technical and practical discussion and development in collaboration with international experts.

Given the extent to which other inequality measures – e.g., effective access and coverage inequality – draw on the conceptual framework of the approach to health inequality (inter-individual inequality in the probability of an event), these measures too should be included in a robust indicator development process that engages appropriate technical and country-based constituencies.

5. References

Braveman, P., B. Starfield, and H. J. Geiger (2001): World Health Report 2000: How it removes equity from the agenda for public health monitoring and policy. *British Medical Journal*, 323(7314): 678-681.

Houweling, T. A., A. E. Kunst, and J. P. Mackenbach (2001): World Health Report 2000: inequality index and socioeconomic inequalities in mortality. *Lancet*, 357(9269): 1671-1672.

Leon, D. A., G. Walt, and L. Gilson (2001): International perspectives on health inequalities and policy. *British Medical Journal*, 322: 591-594.

Szwarcwald, C. L. (2002): On the World Health Organisation's measurement of health inequalities. *Journal of Epidemiology and Community Health*, 56: 177-182.

Ugá A. D., C. M. Almeida, C. L. Szwarcwald, C. Travassos, F. Viacava, J. M. Ribeiro, N. R. Costa, P. M. Buss, and S. Porto (2001): Considerations on methodology used in the World Health Organization 2000 Report. *Cadernos de Saude Publica*, 17(3): 705-712.

Wolfson, M. and G. Rowe (2001): On measuring inequalities in health. *Bulletin of the World Health Organization*, 79(6): 553-560.

IX. RESPONSIVENESS: LEVEL AND DISTRIBUTION

1. WHR 2000

Responsiveness was included as an intrinsic goal in the health systems performance framework because the way people are treated when they come into contact with the system can improve or reduce their well-being independently of health outcomes. There were seven domains of responsiveness: dignity, autonomy, confidentiality of information, prompt attention, access to social-support networks, quality of basic amenities, and choice of health-care provider. For WHR 2000, WHO obtained data on responsiveness through key informant surveys for 35 out of 191 countries. Within each of the 35 countries, a single focal person was canvassed. Each focal person selected an average of 50 key informants from a broad range of health-system stakeholders, including consumer groups, to answer a short questionnaire. Focal persons oversaw data capture and submission of data to WHO. Data from two such surveys was not of sufficient quality to be used in WHR 2000.

For the overall measure of responsiveness, key informants were asked to provide a general rating of the health system in their country with respect to the seven domains after they had answered specific questions on each domain. The specific questions were used to ensure that key informants correctly identified the various components of the domains. Correlation and exploratory factor analysis were undertaken to check for consistency. There was a high degree of consistency between the responses to the specific domain questions and the general rating questions.

In the final analysis, the overall domain scores were regressed on a set of covariates for the 33 surveyed countries, and from these regressions the missing data for the remaining 158 countries were imputed. For the distribution of responsiveness, key informants were asked to identify marginalized groups. This information was used together with the information on the size of those groups in the country to develop a responsiveness inequality score (distribution). Once again, the information was imputed for the 158 non-surveyed countries.

2. Main commentaries and criticisms

Data sources

Many criticisms were raised on this method of obtaining information (Almeida et al. 2001; Navarro 2001; Blendon et al. 2001a) and estimating the missing data (Williams 2000; Almeida et al. 2001; Aalto 2000). Criticisms included the fact that the method was biased because most of the key informants were WHO people; that the method was inherently flawed as it was not a representative sample of the population; that only seven questions out of 42 were used for the index; and that too many imputations were made from too little data. In particular, it was noted that the data and methods used to estimate responsiveness inequality for the unsurveyed countries resulted in multiple tied ranks (Williams 2000).

Relative weight of responsiveness

Commentators from regional consultations questioned the relative importance of having responsiveness in the framework (WHO Regional Office for the Eastern Mediterranean 2001).

Domain weights

Several commentators questioned the relative weights of the seven domains in the aggregation for an index of overall responsiveness (WHO Regional Office for Europe 2001; WHO Regional Office for South-East Asia 2001; WHO Regional Office for the Western Pacific 2001; WHO Regional Office for Africa 2001).

Responsiveness of the broader health system

Comments were made in regional consultations (WHO Regional Office for Europe 2001; WHO Regional Office for South-East Asia 2001) that responsiveness needed to reflect the broader boundaries of the health system, including public access to information and other services of health protection and promotion (see also Ugá et al. 2001; Travassos 2001; Oswaldo Cruz Foundation 2000).

Sources of information

The Blendon et al. (2001a) critique addresses the issue of who is better qualified to judge health-care systems – key informants or users of the system. Blendon et al. (2001b) state that both satisfaction and responsiveness measures are important when information is canvassed from the population.

Translation, validity and reliability

The critique of Aalto (2000) covers questions related to translating the concept of responsiveness and cross-cultural validity. The issue of cross-

cultural validation was also raised in several regional consultations (WHO Regional Office for Europe 2001; WHO Regional Office for South-East Asia 2001; WHO Regional Office for the Western Pacific 2001). Participants in the regional consultations (WHO Regional Office for Africa 2001; WHO Regional Office for the Western Pacific 2001) felt that translation might be a slightly more difficult problem for responsiveness than for other modules owing to the abstractness of the concepts involved (see also Almeida et al. 2001). Aalto (2000) and participants in regional consultations (WHO Regional Office for South-East Asia 2001; WHO Regional Office for the Western Pacific 2001) criticized the availability of standard instrument psychometric data on the responsiveness key informant instrument. Aalto (2000) and the SEARO and WPRO regional consultations (WHO Regional Office for South-East Asia 2001; WHO Regional Office for the Western Pacific 2001) indicated that this type of data should be available for any subsequent responsiveness questionnaire instruments (e.g., in household surveys).

Universality of domains

Aalto (2000) commented extensively on the need to provide a convincing rationale for the choice of domains. The change to household surveys was commended but WHO was cautioned that cross-cultural validation of survey questions on domains should be ensured in any future household survey work. (To some extent this is linked to the issue of translation.) At some of the regional consultations, participants raised the issue of relevance of the domains in different cultural settings (WHO Regional Office for Africa 2001; WHO Regional Office for the Western Pacific 2001; WHO Regional Office for South-East Asia 2001). The critique of Williams (2000) also touched on this issue.

Non-users

A commentary of the Brazilian Ministry of Health (Oswaldo Cruz Foundation 2000) criticized the responsiveness work on the grounds that the WHO indicator was limited to measuring the experiences of people who actually use health services.

3. WHO responses and proposals

Data sources

In order to improve data sources, WHO has focused on developing survey instruments to obtain information from households. The number of countries covered by household surveys will be increased substantially. Some 60 countries have already been surveyed through the Multi-Country Survey Study, and the World Health Survey will cover a further 70 countries. Using

the Multi-Country Survey Study data, distributional measures of responsiveness are being developed and tested.

Relative weight of responsiveness

The relative importance of responsiveness within the overall framework is being tackled with new survey questions in the World Health Survey, which are currently being tested.

Domain weights

Since WHR 2000, WHO has launched the Multi-Country Survey Study in which households were asked directly about their relative weights for each domain. In analyzing the data from this Study, WHO has found that they indicate a common set of rankings of domains across countries, and possibly a tendency towards a common set of weights. However, conclusions on the weight structure across countries are limited by the structure of the original question. New questions to elicit weights from respondents for the domains are being tested in the World Health Survey pilots.

Responsiveness of the broader health system

New questions on health promotion and support structures for families looking after ill family members at home are being tested in the World Health Survey pilots.

Satisfaction versus responsiveness

The WHO responsiveness survey module is designed using the latest thinking in the field of patient assessment measurement, based on patients' interactions with the health-care system. Satisfaction remains an interesting measure for other reasons because it solicits people's opinions about the system, rather than their reports of personal interactions with it. More work is being done to test the use of techniques to improve cross-population comparability of results from surveys of people's experiences.

Translation, validity and reliability

Since WHR 2000, the WHO has developed an extensive translation protocol for the Multi-Country Survey Study, which has been improved further in the piloting of the World Health Survey instrument. In addition, facility surveys are being developed to collect evidence on the validity of the questionnaire instrument. These surveys will enable the comparison of observations on certain domains of responsiveness in facilities with reports from individuals using those facilities. Other standard validity strategies recommended by Aalto (2000) and mentioned by participants in regional consultations (WHO Regional Office for South-East Asia 2001; WHO Regional Office for the

Western Pacific 2001), such as comparisons with similar data series, are being pursued. WHO is also continuing to document the results of standard psychometric tests of the household survey instruments.

Universality of domains

WHO has produced a paper documenting the criteria for selection of the domains (De Silva 2000). Since WHR 2000, an eighth domain of responsiveness – clarity of communication – has been included. Questions relating to this new domain were developed in response to consultations and included in the Multi-Country Survey Study. A group of ethicists was asked to review the cross-cultural dimensions of the responsiveness domains. Their findings were submitted to the technical consultation on responsiveness (World Health Organization 2001). In addition, further work is currently underway to map the responsiveness domains to UN and other international conventions and treaties on human rights. More cognitive testing is planned for the responsiveness module items in the World Health Survey.

Non-users

With respect to this critique, efforts have concentrated on finding ways to include non-user and low-user groups. As a first attempt at addressing the non-user and low-user problem, models to predict responsiveness for non-users and low-users were developed. This proposal was discussed at the technical consultation on responsiveness (World Health Organization 2001), and goes some way to addressing the problem of non-users and low-users. Both the Multi-Country Survey Study and the piloted World Health Survey instruments have included questions regarding utilization.

4. SPRG comments and recommendations

Data sources

SPRG members agree with the criticisms made by external commentators about the data sources. In particular, SPRG members concur that people using health systems should be asked their opinion about it, rather than relying on information from key informants. They recommend that if the indicator of responsiveness is to be utilized in future, it will be necessary for WHO to obtain representative household-level data for all countries.

Relative weight of responsiveness

Some SPRG members were concerned with the inclusion of 'responsiveness' for the evaluation of health-systems performance. Responsiveness as defined by WHO is meant to deal with the interactions of users with the health system, and includes features such as respectful treatment,

confidentiality of information, prompt attention, and involvement in decision-making. Such features apply to many service activities, e.g., educational services, transportation services, etc. Some SPRG members therefore felt that it might be better to deal with issues of responsiveness generically, i.e., at the national (or even international) level rather than at the health-system level. Other members rejected this notion and pointed out that if responsiveness was measured at the national level, accountability could only be attributable at that level – and not at the health-system level.

SPRG members were also surprised that the weight on responsiveness in the composite indicator was as high as 25%, the same weight as for average health level. This implies that a one-point increase in responsiveness is valued as highly as a one-point increase in the scale used for health (equivalent to almost one year of health-adjusted life expectancy). In view of the implied trade-off, SPRG members wondered whether appropriate questions had been asked to elicit the relative weights for responsiveness and the intrinsic goal of average health level. SPRG members also wondered whether it might have been appropriate to incorporate a changing set of weights for responsiveness at different stages of health-system development. (The present set of weights is constant for all levels of development.) It could be argued that a greater weight attaches to pure health goals relative to responsiveness at low levels of life expectancy (e.g., 50 years) than at high levels of life expectancy. Once life expectancy reaches 70 or more years, as in the OECD countries, it may be more appropriate to use a relatively larger weight on responsiveness. At high levels of life expectancy the room for further improvement in health is limited, and other goals – such as responsiveness of the health system – may assume greater importance.

Domain weights

Several commentators (WHO Regional Office for Europe 2001; WHO Regional Office for South-East Asia 2001) suggested that the responsiveness domains should be given country-specific weightings in the aggregation of the domains into an index of overall responsiveness (while maintaining the relative weight of overall responsiveness *vis-à-vis* other goals of the health system). SPRG members also recommend that WHO experiments with a non-linear system of weights to reflect changing priorities that might attach to responsiveness relative to pure health goals at different stages of development. It was recognized that some of the responsiveness measures deal with human rights issues, such as dignity and confidentiality, which need to be addressed at all stages of development. An appropriately specified non-linear system of weights can accommodate constant linear weights on certain domains of responsiveness.

Some SPRG members suggested that WHO assess the relationship between the level of responsiveness (by domain) in a country and the level of financial resources available to its health system. This approach will help assess

whether there is a differential capacity in countries for producing responsiveness.

Responsiveness of the broader health system

Some SPRG members also questioned the use of the term 'responsiveness'. Responsiveness of the health system could be construed to include several other features apart from interactions with the population – such as the delivery of health services, health promotion and protection, and health education. The term has often been confused with the notion of how well the health system 'responds'. In consequence, some SPRG members suggested that WHO should consider changing the term 'responsiveness' to something like 'interactions with users'. Other SPRG members suggested that possibly a term like 'patient-/people-centredness' or 'patient rights' might work.

Some SPRG members suggested that WHO should conduct a thorough survey to identify potential questions that address responsiveness as it relates to health promotion and disease prevention.

Based on the critiques and their own assessment, SPRG felt that the present WHO questionnaire on responsiveness was geared to eliciting information mainly on personal health services, and that health promotion and protection activities were relatively neglected. Some SPRG members also wished to see the responsiveness of financing activities assessed.

Taken together, SPRG felt that there was a case for extending the responsiveness domains to aspects of health-system activities beyond personal health-care services – e.g., early warning systems in the case of epidemics or other biological or environmental health threats, health promotion and protection, health education in schools, research, etc.

Satisfaction versus responsiveness

SPRG agrees with Blendon et al. (2001a) that users rather than key informants should be the judges of the health-care system. SPRG acknowledges the usefulness of satisfaction measures in general, but feels they are not necessarily a substitute for responsiveness in the framework of HSPA. For example, a person might feel satisfied because he was cured, but he may not have received prompt attention or have been treated with respect. Alternatively, a person might feel satisfied if he were prescribed drugs, even if these drugs were unnecessary, or harmful, for his condition.

In this regard it should be noted that, unlike responsiveness, measures of satisfaction do not adjust for people's differing expectations of the health system. This adjustment is made through the HOPIT approach (see Section XIII on Cross-Population Comparability).

SPRG recommends that WHO should continue work on developing experience measures and the use of vignettes and other techniques for dealing with cultural differences in expectations and response tendencies.

Translation, validity and reliability

Ensuring the accuracy of translation is a difficulty faced by all surveys administered in multiple languages. In particular, in a country with many dialects, the infeasibility of issuing a questionnaire in all its languages and dialects presents obvious problems in administering interviews. (For example: Was the interviewer able to communicate in the respondent's dialect? How well was the interviewer able to translate concepts and questions on the spot? Did he use exactly the same wording for different households?) SPRG members as well as participants in the regional consultations (WHO Regional Office for Africa 2001; WHO Regional Office for the Western Pacific 2001) felt that translation might be a slightly more difficult problem for responsiveness than for other modules owing to the abstractness of the concepts involved.

SPRG recommends that WHO conducts more extensive cognitive testing to evaluate how respondents interpret the survey items. In addition, rigorous interviewer training protocols need to be developed, tested and applied. Training and management of interviews must meet high standards to try to ensure the consistent application of the survey protocols.

Universality of domains

SPRG recommends that WHO should document the mapping of cultural influences on responsiveness domains as well as the mapping of domains onto UN and other international conventions and treaties on human rights.

Non-users

In noting the Brazilian comments (Oswaldo Cruz Foundation 2000, Travassos 2001), SPRG felt that there were indeed serious problems in using an indicator that was limited to measuring the experiences only of people who use health services, especially when making cross-country comparisons. For example, it could turn out that only 20% of the population of country A used its health-care system, and this system was judged to be perfectly responsive by its users (according to the scoring criteria). In contrast, in country B, 80% of the population used its health-care system, which was judged to be only 50% responsive by its users. Which system is more responsive?

According to the WHR 2000 definition of responsiveness (experience of users), the health-care system of country A is more responsive. Several SPRG members expressed unease with this logical conclusion. However, the conclusion is inevitable if coverage is not a part of the definition of

responsiveness. Indeed, according to some SPRG members, the term 'responsiveness' evokes the idea of a health-care system responding to people's needs. Hence, if people in country A have been put off from using the system (because of out-of-pocket costs, lack of knowledge, high transport costs, previous bad experience, etc.), this should be reflected in any measure of the responsiveness of the system.

SPRG noted the development of an approach by WHO to predict responsiveness among non-users. However, if the WHO maintains its current approach to measuring responsiveness among the actual users of the system, SPRG recommends that measures of responsiveness should be accompanied by measures of utilization.

5. References

Aalto, A. M. (2000): Measuring the responsiveness of health care system in the World Health Report 2000. In eds. Häkkinen, U. and Ollila, E. *The World Health Report 2000: What does it tell us about health systems? Analyses by Finnish Experts*. Helsinki, Finland: National Research and Development Center for Welfare and Health (STAKES).
[<http://www.stakes.fi/english/publicati/Publications.htm>]

Almeida, C., P. Braveman, M. R. Gold, C. L. Szwarcwald, J. M. Ribeiro, A. Miglionico, J. S. Millar, S. Porto, N. R. Costa, V. O. Rubio, M. Segall, B. Starfield, C. Travassos, A. Ugá, J. Valente, and F. Viacava (2001): Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357(9269): 1692-1697.

Bernard, R. H. (1994): *Research Methods in Anthropology. Qualitative and Quantitative Approaches*. AltaMira Press.

Blendon, R. J., M. Kim, and J. M. Benson (2001a): The public versus the World Health Organization on health system performance. *Health Affairs*, 20(3): 10-20.

Blendon, R. J., M. Kim, and J. M. Benson (2001b): Authors respond to WHO critics. *Health Affairs*, 20(4): 253.

Darby, C., N. Valentine, and C. J. L. Murray (2000): WHO strategy on measuring responsiveness. Global Programme on Evidence for Health Policy Discussion Papers, No. 23. Geneva, Switzerland: World Health Organization.

De Silva, A. (2000): A framework for measuring responsiveness. Global Programme on Evidence for Health Policy Discussion Papers, No. 32. Geneva, Switzerland: World Health Organization.

De Silva, A. and N. Valentine (2000b): Measuring responsiveness: results of a key informant survey in 35 countries. Global Programme on Evidence for Health Policy Discussion Papers, No. 21. Geneva, Switzerland: World Health Organization.

Häkkinen, U. and E. Ollila, eds. (2000): *The World Health Report 2000: What does it tell us about health systems? Analyses by Finnish experts*. Helsinki: National Research and Development Center for Welfare and Health (STAKES).
[<http://www.stakes.fi/english/publicati/Publications.htm>]

Murray, C. J. L., J. Frenk, D. Evans, K. Kawabata, A. Lopez, and O. Adams (2001): Science or marketing at WHO? A response to Williams. *Health Economics*, 10(4): 277-282.

- Murray, C. J. L., K. Kawabata, and N. Valentine (2001): People's experience versus people's expectations. *Health Affairs*, 20(3): 21-24.
- Navarro, V. (2002): The World Health Report 2000: Can health care systems be compared using a single measure of performance? *American Journal of Public Health*, 92(1): 31-34.
- Navarro, V. (2001): World Health Report 2000: Response to Murray and Frenk. *Lancet*, 357(9269): 1701-1702.
- Navarro, V. (2000): Assessment of the World Health Report 2000. *Lancet*, 356(9241): 1598-1601.
- Oswaldo Cruz Foundation, Brazilian Ministry of Health (2000): Report of the workshop "Health Systems Performance: The World Health Report 2000". 14-12-2000, Rio de Janeiro.
- Travassos, C. (2001): Assessing health systems performance -- a critical appraisal about the WHO World Health Report 2000 and future developments. Paper presented at Conference on Restructuring of Health Services and Corporate Public Health in the Era of Reforms. 5-7-2001, Maastricht, The Netherlands.
- Ugá, A. D., C. M. Almeida, C. L. Szwarzwald, C. Travassos, F. Viacava, J. M. Ribeiro, N. R. Costa, P. M. Buss, and S. Porto (2001): Considerations on methodology used in the World Health Organization 2000 Report. *Cadernos de Saude Publica*, 17(3): 705-712.
- Ustün B., S. Chatterji, M. Villanueva, L. Bendib, C. Celik, R. Sadana, N. Valentine, C. Mathers, J. P. Ortiz, A. Tandon, J. Salomon, Y. Cao, X. W. Jun, and C. J. L. Murray (2000): WHO multicountry household survey study on health and responsiveness 2000-2001. Global Programme on Evidence for Health Policy Discussion Papers, No. 37. Geneva, Switzerland: World Health Organization.
- Ustün, B. and C. J. L. Murray on behalf of the World Health Organization Survey Programme (2001): World Health Survey: objectives, design, modules. Global Programme on Evidence for Health Policy, mimeo. Geneva, Switzerland: World Health Organization.
- Valentine, N., J. P. Ortiz, A. Tandon, K. Kawabata, and C. J. L. Murray (2001): Health System Responsiveness: Evidence from Population Surveys in 15 Countries. Global Programme on Evidence for Health Policy, mimeo. Geneva, Switzerland: World Health Organization.
- Valentine, N., J. Salomon, and C. J. L. Murray (2001): Weights for responsiveness domains: analysis of country variation in 57 national sample

surveys. Global Programme on Evidence for Health Policy, mimeo. Geneva, Switzerland: World Health Organization.

Williams, A. (2000): Science or marketing at WHO? A Commentary on 'World Health 2000'. *Health Economics*, 10(2): 93-100.

World Health Organization (2001): Report on WHO meeting of experts: Responsiveness. 13-9-2001, Geneva, Switzerland.

WHO Regional Office for Africa (2001): General Report on the Regional Consultative Meeting on Health Systems Performance Assessment: Final Report. 18-5-2001, Harare, Zimbabwe.

WHO Regional Office for the Americas (2001): Regional consultation of the Americas on Health Systems Performance Assessment, Background Document: Critical issues in health systems performance assessment. Washington, D.C.

WHO Regional Office for the Eastern Mediterranean (2001): Report on the Regional Consultation on the Conceptual Framework for Health System Performance Assessment. 9-7-2001, Ain Saadeh, Lebanon.

WHO Regional Office for Europe (2001): Report of the Regional Consultation on Health Systems Performance Assessment. 3-9-2001, Copenhagen.

WHO Regional Office for South-East Asia (2001): Report of the regional consultation and technical workshop on health systems performance assessment. 18-6-2001, New Delhi, India.

WHO Regional Office for the Western Pacific (2001): Report of the regional consultation on health system performance. 3-7-2001, Manila, Philippines.

X. FAIRNESS IN FINANCIAL CONTRIBUTION

1. WHR 2000

WHO began its analysis of the fairness of household financial contributions to a health system by asking the question: Taking society's efforts to redistribute income as a given, what are fair contributions by households to the health system? As a normative claim, WHO proposed that the sacrifice created by contributing to the health system should be equalized across households independent of their health status or their utilization of health services. This 'equal sacrifice' was interpreted as an equal share of each household's capacity to pay. The goal of the health system was not seen to involve the redistribution of income, but was based on the notion that a health system should 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 taken to be the level of consumption of the household (or 'permanent income' in a life-cycle perspective). Subsistence expenditure typically includes spending on food, basic shelter and minimal clothing, but not on health. However, in WHR 2000, it was not possible to obtain estimates of spending for basic shelter and minimal clothing, so capacity-to-pay was measured simply as total consumption expenditure minus food consumption, and a household's financial contribution (HFC_i) was measured as its total payments to the health system divided by its capacity-to-pay (CTP_i).

A fairness of financial contribution index was constructed to measure inequality in the distribution of household financial contributions. As catastrophic spending was considered to be the first concern of the health-financing system, WHO argued that households with catastrophic health expenditure should be given more weight in the index of inequality. The index should take into account catastrophic expenditure (extreme horizontal inequality) as well as moderate horizontal inequality, and incorporate people's expectations that the rich should pay more in absolute terms than the poor.

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, and was built on the assumption that people care about their place in relation to the average contribution and judge inequality accordingly. The alternative would be that people care about their place relative to other individuals, and not to the average. The choice of the former index was made after conducting an internet survey of over 1,600 people in which a majority of people appeared to care more about the difference between their contribution and the average for the population than the difference between their contribution and that of every other individual in the population.

The index was estimated for 21 countries which had recently conducted household income and expenditure surveys. The results for these countries were used in a regression analysis to identify critical covariates of the FFC index, and this regression was used to estimate FFC for the remaining 170 countries.

2. Main commentaries and criticisms

The commentaries and criticisms covered three general areas: technical questions about the index used in WHR 2000; problems with data availability; and the policy relevance of the FFC index.

The FFC index

The first technical question raised about the FFC index used in WHR 2000 was that it could penalize countries with very progressive payment systems because perfect fairness was defined as the situation in which all households contribute an equal share of their capacity-to-pay. Countries with very progressive payments where the rich pay a higher proportion of their capacity-to-pay than the poor would then be shown as having an unfair health system (Shaw 2001; Wagstaff 2001; Ministry of Health, Vietnam 2001; Wagstaff and Van Doorslaer 2001; Travassos 2001; Ugá et al. 2001; WHO Regional Office for the Americas 2001; World Health Organization 2001).

Secondly, the FFC Index is found to be relatively insensitive to vertical inequality (Ammar and Kasperian 2001), an aspect that policy makers can target – as opposed to horizontal inequality, which they cannot so easily target.

Thirdly, measuring capacity-to-pay as total expenditure minus food consumption was criticized because much of the food consumption of the rich is not subsistence spending (Klavus 2000; Navarro 2000). Subtracting food consumption from total expenditure may underestimate the capacity-to-pay of rich households (Ammar and Kasperian 2001).

Fourthly, the technical consultation and some of the regional consultations suggested that the interval-scaling properties of the FFC Index had not been established and that the units of the index were not interpretable (Szwarcwald 2002).

Data availability

Household income and expenditure surveys that had appropriate information for the construction of the index were available for only 21 countries. This was not seen as sufficient for estimating an FFC Index for all 191 Member States – there was simply too much missing data (Williams 2000; Nord 2002).

Policy relevance of the FFC index

Some commentators at the regional consultations suggested that policy makers needed to have the ability to drill down to the components of the index and to identify the impact on vertical inequality, horizontal inequality and catastrophic payments separately. A second issue was whether policy makers might be interested in the income redistributive effects of health payments in addition to inequality in the financial burden of payments (Wagstaff 2001).

Thirdly, the FFC Index does not take account of the utilization of health services (Travassos 2001). A system in which all people pay the same low proportion of their capacity-to-pay would be fair according to the index, but would give no indication about whether people were unable to obtain the care they needed because of its cost.

3. WHO responses and proposals

Since publication of WHR 2000, WHO has undertaken a considerable body of analytical work to explore the implications of the commentaries, criticisms and suggestions made at the consultations, in the literature, and by SPRG members. This has led to a number of background documents being prepared for SPRG and the following proposals have been put to SPRG for the next round of performance assessment.

WHO proposes to report routinely on four indicators of the fairness of financial contributions. The FFC Index focuses on the impact of payments on a household's financial burden – in what is referred to as the 'burden space'. In addition to the FFC Index, which summarizes inequality in the distribution of financial burdens, a threshold measure – the proportion of households facing catastrophic expenditures due to health payments – will also be reported. Because of interest expressed by policy makers and researchers since WHR 2000, WHO also proposes to estimate and report the impact of health payments in the 'income space'. Two indicators would be used – the impact of health payments on the overall income distribution, and the percentage of households who fall below the poverty line due to health payments. WHO background documents prepared for SPRG showed that it is feasible to use both indicators with the available data, although the different indicators do attempt to capture different concepts.

For the FFC Index, WHO proposes to change the mathematical formula by using a cube root function to transform the index back into natural units, thereby improving its interpretability.

$$FFC = 1 - \sqrt[3]{\frac{\sum_{i=1}^n |HFC_i - HFC_o|^3}{n}} \quad \text{where } HFC_o = \frac{\sum HE_i}{\sum CTP_i}$$

This index still belongs to the individual-mean family and it is an absolute measure of inequality. It retains the property of the earlier FFC index in that it places a larger weight on households with catastrophic expenditure.

WHO also proposes to change the measure of household capacity-to-pay (CTP) in response to the criticism that food expenditure is not a good indicator of subsistence expenditure. CTP will be redefined as total household expenditure minus the level of expenditure corresponding to the international poverty line (in local currency), as long as total expenditure exceeds this poverty line. In households where total expenditure is less than the estimated poverty line, CTP will be taken to be total expenditure minus the actual food expenditure of the household. The poverty line estimate of 'subsistence' is lower for rich households than the total food expenditure that was used in WHR 2000 to define 'subsistence', which increases the apparent capacity-to-pay of the rich.

Perfect fairness is still defined as each household contributing an equal share of its (redefined) CTP. Although it is theoretically possible that countries with very progressive tax systems may depart from total fairness according to this definition, preliminary empirical results from 55 countries suggest that this does not happen in practice.

WHO also reported to SPRG the results of decomposition of the FFC Index into different components – those due to extreme horizontal inequality related to catastrophic health expenditure, to mild horizontal inequality, and to vertical inequality. Vertical inequality has a small measured component relative to horizontal inequality. For countries with a low value of the FFC Index, inequality is primarily attributable to household catastrophic spending.

Since the publication of WHR 2000, intensive efforts have been made in collaboration with countries to identify new data sources. Currently 104 surveys from 80 countries are available and WHO proposes to continue to seek new sources of data. This will be in addition to questions included in the World Health Survey on assets.

Finally, it is possible that two health systems have the same FFC score – for example, in one system everyone can afford health services, but in the other system a part of the population cannot. WHO proposes, however, that in the second case the population will show poorer levels of health and greater inequalities in health, *ceteris paribus*, than in the first. Hence the problem of poor access will be reflected in poorer health outcomes and in lower overall goal attainment. To try to account for non-use of services because of inability to pay would be double counting in the FFC Index.

4. SPRG comments and recommendations

SPRG endorses the suggestion of routinely reporting on four types of measures of the impact of household financial payments – two in the burden space, and two in the income space. This provides information that is useful to policy makers for different questions that they might wish to address. SPRG also accepts that the new mathematical formulation of the FFC Index is an improvement on the original formulation. The need to obtain household survey data from many more datasets was a common and valid source of criticism of WHR 2000. SPRG emphasizes the need for WHO to reduce the estimation of 'missing data' to the minimum.

SPRG members noted that the cubing formula for the FFC Index in WHR 2000 may have been responsible in large measure for the finding that horizontal inequality accounted for most of the inequality in financial burdens. Another factor responsible for the relatively small component of vertical inequality (compared with horizontal inequality) was that progressivity had already been built into the index through the definition of capacity-to-pay. SPRG members hypothesized that the greater the degree of progressivity that is built into capacity-to-pay, the smaller will be vertical inequality relative to horizontal inequality in the decomposition of the FFC Index.

Some members were concerned about non-utilization of health services by the poor because it was unaffordable. This would lead to the poor making zero financial contributions compared with rich users making significant contributions, and result in an overestimate of measured progressivity (Ammar and Kasperian 2001).

Although the WHO proposal to take the cube root of the differences in the original formula (to transform it back into natural units) might yield better interval-scale properties for the FFC Index, it would make decomposition into various components more difficult to undertake. The decomposition of the FFC Index would no longer be additively separable into identifiable and easily interpretable components of inequality.

SPRG has the following further comments and recommendations.

- (i) WHO should explore ways of controlling for differences in reference periods over which households are asked to report their expenditures. For any given pattern of health expenditures and income flows, the financial burden (or ratio) will be very sensitive to the time frame over which expenditures in the numerator and denominator are measured. An expenditure that is deemed to be catastrophic for a one-week reference period may not be considered catastrophic over a one-month (or longer) reference period. As reference periods differ among the country surveys that are used for the analysis of fairness of financial contributions, cross-country comparisons cannot be made without controlling for these differences.
- (ii) Independently of the empirical problems involved in comparing the incidence of catastrophic health expenditures and inequality in FFC across countries, there is a prior conceptual question as to the appropriate time period for assessing financial burden. WHO needs to elaborate and justify the concept of burden with respect to which the fairness of financial contributions should be assessed.
- (iii) SPRG supports the use of the poverty line to define capacity-to-pay, and encourages WHO to explore the use of variable poverty lines across different regions.
- (iv) The burden need not be defined simply in terms of capacity-to-pay measured as the ratio of expenditure flows over the appropriate time period. An alternative definition might include stock variables in the denominator such as financial and other assets. WHO should explore ways – methodological and empirical – of introducing household assets into the calculation of capacity-to-pay.
- (v) In different health insurance systems, there are differences in time-lags between incurring a health expenditure and receiving

reimbursement. This can affect the comparability of FFC across different settings.

- (vi) Some SPRG members noted that inequality in out-of-pocket payments has different policy implications compared to inequality in overall health financing. They also wished to see an assessment of the financial barriers to fair usage of health services.
- (vii) Inequality in financial contributions is affected by utilization of health services when these are paid for out-of-pocket. In predominantly private health-care systems, the poor may not use services because they cannot afford them. WHO should explore the biases that result from comparing measured inequality of financial burdens when there are different degrees of use and non-use of the system. The present WHO measure of FFC compares systems where financial contributions and utilization are independent with systems where they are endogenous and one depends on the other.

5. References

Almeida, C. M., P. Braveman, M. R. Gold, C. L. Szwarcwald, J. M. Ribeiro, A. Miglionico, J. S. Millar, S. Porto, N. R. Costa, V. O. Rubio, M. Segall, B. Starfield, C. Travassos, A. Ugá, J. Valente, and F. Viacava (2001): Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357(9269): 1692-1697.

Ammar, W. and R. Kasperian (2001): What is fair in financing fairness? *Lebanese Medical Journal*, 49(3).

Klavus, J. (2000): The Measure of Fair Financing in the World Health Report 2000. In eds. U. Häkkinen and E. Ollila. *The World Health Report 2000: What does it tell us about health systems? Analyses by Finnish Experts*. Helsinki: National Research and Development Center for Welfare and Health (STAKES). [<http://www.stakes.fi/english/publicati/Publications.htm>]

Ministry of Health, Vietnam (2001): Comments and suggestions of Vietnam Ministry of Health/Health Policy Unit as regards the World Health Report 2000.

Navarro, V. (2000): Assessment of the World Health Report 2000. *Lancet*, 356(9241): 1598-1601.

Nord, E. (2002): Measures of goal attainment and performance: A brief, critical consumer guide. *Health Policy*, 59(3): 183-191.

Shaw, R. P. (2001): The conceptual basis and the scope of health systems performance assessment in the World Health Report 2000. Invited commentary: Regional consultation of the Americas on health systems performance assessment. 8-5-2001, Washington, D.C. [http://w3.whosea.org/hspa/back_cont_commentry.htm]

Szwarcwald, C. L. (2002): On the World Health Organisation's measurement of health inequalities. *Journal of Epidemiology and Community Health*, 56: 177-182.

Travassos, C. (2001): Assessing health systems performance -- a critical appraisal about the WHO World Health Report 2000 and future developments. Paper presented at Conference on Restructuring of Health Services and Corporate Public Health in the Era of Reforms. 5-7-2001, Maastricht, The Netherlands.

Ugá, A. D., C. L. Szwarcwald, C. Almeida, C. Travassos, F. Viacava, J. M. Ribeiro, N. R. Costa, P. Buss, and S. Porto (2000): Views on the WHO 2000 Report. Rio de Janeiro: FIOCRUZ.

Ugá, A. D., C. M. Almeida, C. L. Szwarcwald, C. Travassos, F. Viacava, J. M. Ribeiro, N. R. Costa, P. M. Buss, and S. Porto (2001): Considerations on methodology used in the World Health Organization 2000 Report. *Cadernos de Saude Publica*, 17(3):705-712.

Wagstaff, A. (2001): Measuring equity in health care financing: Reflections on and alternatives to the World Health Organization's Fairness of Financing Index. Development Research Group and Human Development Network, World Bank. [<http://www.healthsystemsrc.org/>]

Wagstaff, A. and E. Van Doorslaer (2001): Paying for health care: Quantifying fairness, catastrophe, and impoverishment, with applications to Vietnam 1993-1998. Paper presented at International Health Economics Association (iHEA) Conference. July 2001, York, United Kingdom.

Williams, A. (2000): Science or marketing at WHO? A Commentary on 'World Health 2000'. *Health Economics*, 10(2): 93-100.

WHO Regional Office for the Americas (2001): Regional consultation of the Americas on Health Systems Performance Assessment: Final Report. 8-5-2001, Washington, D.C.

World Health Organization (2001): Report on WHO technical consultation on fairness in financial contribution. 4-10-2001, Geneva, Switzerland.

XI. COMPOSITE GOAL ATTAINMENT

1. WHR 2000

A composite index of goal attainment was constructed for each Member State as a weighted sum of attainment on each intrinsic goal (Gakidou et al. 2000; Murray et al. 2000). Weights were obtained from a world-wide-web key informant survey involving more than 1,600 participants from over 100 countries. Fifty percent of the total weight was ascribed to health (25% to the average level and 25% to inequality), 25% to fairness of financial contributions, and 25% to responsiveness (12.5% to the average level and 12.5% to inequality). Uncertainty intervals were reported for the scores on the attainment index and the associated ranks.

2. Main commentaries and criticisms

The question of the composite indicator has perhaps received more comments and criticisms than any other issue related to WHR 2000. Some comments have been favourable: for example, the African regional consultation suggested that a composite index may be useful for comparison purposes (WHO Regional Office for Africa 2001). The Americas regional consultation indicated that the direct comparison offered by the composite index may help ministries of health secure increased political attention (WHO Regional Office for the Americas 2001). Several consultations indicated that the composite index might be useful for comparing health systems in countries with similar economic and other background characteristics. Some felt that the index could become useful in the future if the underlying science were improved (WHO Regional Office for South-East Asia 2001).

Many contrary opinions have also been expressed (for example, Ugá et al. 2001; Hurst and Jee-Hughes 2001; Almeida et al. 2001). These arguments can be considered under two broad headings: objections in principle and scientific objections. The objections in principle can be summarized as follows.

- nations have different objectives and priorities with respect to their health systems, which a single composite index cannot capture (Navarro 2000 and 2002; Ozwaldo Cruz Foundation 2000);
- nations operate in different environmental, economic and political circumstances, and comparison is either inappropriate or infeasible (Nord 2002; Häkkinen 2000);
- the composite is not helpful as it offers no policy guidance – more disaggregate data are needed (Nord 2002);

- many countries do not have the capacity to interpret the implications of the index, and so may make inappropriate policy responses;
- the rankings implicit in the composite index generate media coverage that may be unhelpful or misleading (Lancet 2001).

The scientific objections that were made about the WHR 2000 composite index can be summarized as follows:

- there was no agreement on whether the five components of the index were universally appropriate (Coyne and Hilsenrath 2002);
- the components of the index refer to different definitions of the health system (for example, health outcomes to a very broad definition, responsiveness to a narrow definition based predominantly on health care);
- the components of the index refer to different time periods (for example, health outcomes to a long period, responsiveness to the current period);
- the rescaling of the component indicators onto a 0 to 100 scale was arbitrary, and its consequences difficult to understand;
- the weights used in the composite index were derived from key informant interviews and were not representative of population preferences (Almeida et al. 2000; Williams 2000; Smith 2002);
- the methodology for deriving the weights was flawed – in particular, the questionnaire used did not elicit the required relative *marginal* valuation of an extra unit of performance (Smith 2002);
- the measurement of the individual components of the index was poor;
- the treatment of 'missing data' was inadequate, and there were too many missing data to make the composite indicator credible (Nord 2002; McKee 2001; Häkkinen 2000);
- the rankings reported in WHR 2000 are sensitive to the weights used (Oswaldo Cruz Foundation 2000);
- the methods used were not validated or exposed to adequate scientific review.

Specific recommendations in the literature included:

- WHO should publish the underlying data, but not aggregate it into a single index;
- comparisons should be reported only for clusters of comparable countries;
- different transformations (such as z-scores) should be used for the component measures (Oswaldo Cruz Foundation 2000);
- different weights or component measures might be used for different clusters of countries, reflecting different circumstances, priorities and objectives;
- WHO should offer more support for understanding the composite scores and translating into local action;
- satisfaction, coverage and process measures should be incorporated into the index;
- better methodology should be adopted for inferring weights (Appleby and Street 2001);

- better methodology should be adopted for the analysis of uncertainty;
- a research and development effort on the use on composites should be considered by WHO.

Many different suggestions were made about the advisability of continuing to publish a composite index. Some participants in the regional consultations felt that the Human Development Index had played a useful role in mobilizing opinion and political commitment, and that an aggregate index of health-system attainment could play a similar role. At the other extreme, critics felt that WHO should publish the underlying data on attainment of individual goals but should not aggregate the scores into a single index (Nord 2002).

Other commentators felt that WHO should make comparisons only within clusters of comparable countries (rather than among all 191 Member States taken together), and that it would be appropriate to use different sets of weights or goals for different clusters of countries (Nord 2002). If WHO chose to continue with a composite attainment index, it should offer more support to countries to understand its meaning and to translate the results of the exercise into better policy.

3. WHO responses and proposals

WHO has examined some of the above criticisms in preparing for the SPRG meetings. For example, the variability of weights was explored from representative population samples in more than 60 countries as part of the WHO Multi-Country Survey Study 2000-2001. Although the weights do vary, in no household survey was the average reported weight equal to zero for any component. SPRG was presented with scores from 53 countries for which data have been analyzed ($n > 51,000$) and the average weights were 46% for health (25% for average level and 21% for inequality), 26% for fairness of financial contributions, and 28% for responsiveness (15% for average level and 13% for inequality). These weights are similar to the weights obtained from the internet survey conducted in 2000.

WHO also examined two methods of recalculating the overall attainment index reported in WHR 2000 to take account of differences in weights observed in the Multi-Country Survey Study (Lauer et al. 2002). In the first, each country's weights were allowed to vary between the minimum and maximum weights observed across all survey countries. For each country, weights were chosen from within this range so as to maximize its overall attainment score, given the country's scores on the five separate goals. This procedure resulted in the highest overall attainment score for each country (with the weights constrained to lie in the ranges observed), and was termed the 'benefit-of-the-doubt' score. The rank correlation between the WHR 2000 score and the 'benefit-of-the-doubt' score of countries was 0.997.

In the second method, weights were again constrained to lie within the ranges observed across the survey countries, but mathematical programming techniques (data envelopment analysis) used in Operations Research were applied to determine the best weights for each country. Weights calculated in this way yielded a third overall attainment score. The rank correlation between this alternative 'benefit-of-the-doubt' score for countries and the original WHR 2000 score was 0.984. The ranking of countries changed as a result of using these two alternative types of 'benefit-of-the-doubt' weights, but all ranks remained within the uncertainty intervals reported in WHR 2000.

WHO has therefore proposed to SPRG that the composite index should continue to be calculated and reported on routinely. Those who prefer to focus on the individual goals can do so because the separate scores would still be reported. To provide a basis for comparability, the average weight across countries would be used to estimate the overall attainment index. Overall attainment using 'benefit-of-the-doubt' weights would also be reported, as would an index based on the weights estimated for each country. Finally, WHO will continue to investigate whether there are systematic determinants of country weights, and will explore alternative methods of eliciting weights for the goals, including the use of survey questions involving trade-offs.

4. SPRG comments and recommendations

Smith (2002) has examined the case for developing a composite score of health-system performance. In summary, the arguments in favour of developing composite indicators of performance (as distinct from separate consideration of the component indicators) include the following.

- They place system performance at the centre of the policy arena, and draw the attention of senior policy makers to the issue.
- They can offer a rounded assessment of system performance.
- They enable subsequent judgements to be made on system efficiency.
- They facilitate communication with citizens and promote accountability.
- They indicate which systems represent the beacons of best performance.
- They indicate which systems represent a priority for improvement efforts.
- They may stimulate the search for better data and better analytic efforts across all of health care.
- Use of a composite performance measure recognizes the trade-offs that exist between different objectives, and leaves local policy makers free to decide along which indicators they have greatest scope for improvement.

Against this, the use of composite indicators (in preference to piecemeal scrutiny of individual performance measures) can lead to dysfunctional outcomes for the following reasons.

- By aggregating individual measures of performance, composite indicators may disguise serious failings in some parts of some systems.
- As measures of performance become more aggregate, it becomes increasingly difficult to know to what to attribute poor performance, and therefore what remedial action to take.
- The individual elements used in the composite indicator can often be contentious.
- A composite that seeks to be comprehensive in its coverage may have to rely on very feeble or opaque data in some dimensions of performance.
- A composite that ignores dimensions of performance that are difficult to measure may give misleading messages and distort behaviour in undesirable ways.
- Current methodology for the calculation of weights is still inadequate.
- The weights used in composite indicators reflect a single set of preferences. Yet there may exist great diversity in preferences amongst policy makers and ordinary citizens – in short, a composite indicator does not respect alternative viewpoints.

In light of these observations, SPRG considers that the first requirement is that WHO makes a strategic decision whether it wishes to continue with publication of the composite scores and rankings. There will always be variation in the weights attached by individuals and nations to health-system objectives, and the decision to construct a composite is therefore ultimately a strategic (or policy) decision rather than a scientific judgement. However, the practical scientific difficulties of developing a satisfactory composite score may be an important element in informing this strategic decision.

If a decision is taken to continue to publish a composite attainment score, SPRG believes that WHO should indicate clearly that the science of composite indices is still in the process of development. Any results from this analysis should not be interpreted as a definitive judgement on health-system attainment. In addition, the following scientific issues arise.

- (i) The fact that the different components of the composite index relate to different concepts of the 'health system' needs careful attention. For example, it is unrealistic to attribute current health outcomes to the current health system. For this reason, we recommend that the components of the composite index should be reconsidered in the light of the responses to WHR 2000. One possibility would be to examine whether some measures of process should be included.
- (ii) The quality of many of the data used in constructing the WHR 2000 composite index was deficient. We welcome the subsequent efforts made by WHO to improve the quality of the measurement instruments used and the availability of data, and recommend that the process of data improvement continues to be given a high priority.

- (iii) The treatment of 'missing data' in WHR 2000 was inadequately documented. We welcome signals that WHO is beginning to develop its thinking about this issue (Murray et al. 2001). SPRG recommends that WHO methodology in this area is developed further, in discussion with relevant experts, and that the technical judgements made in the treatment of 'missing data' are transparent and well-documented.
- (iv) SPRG welcomes the principle of seeking to report uncertainty intervals around estimates of attainment. The WHR 2000 analysis of uncertainty included the construction of distributions (of estimates of attainment) based on sampling error and parameter estimation. However, it did not include 'second-order' sources of error, such as model specification or measurement errors. We recommend a more transparent approach to the treatment of uncertainty, which may *inter alia* require reconsidering the basis of the 'sampling distributions' from which uncertainty intervals are calculated.
- (v) In order to construct the composite indicator, each of its constituent components is transformed onto a common scale of 0 to 100. These transformations are inextricably linked to the set of weights used in the composite, and should in principle be designed such that the chosen set of weights is valid at every level of attainment. (Alternatively the weights could be allowed to vary depending on the level of attainment.) We therefore recommend that WHO reconsiders the methods it uses to transform indicators, and ensures that they are consistent with the set of weights employed.
- (vi) The derivation of the weights used in the composite index in WHR 2000 was rightly criticized for a number of reasons. It is imperative that WHO reconsiders its methodology for eliciting weights in order that the inferred weights are consistent with the scales used to measure a unit of attainment in each dimension. SPRG welcomes WHO efforts to seek a more representative basis for deriving weights (through WHS) and using more scientific methods to elicit respondents' preferences. We recommend that these efforts are pursued with vigour, in consultation with relevant experts.
- (vii) SPRG believes it is imperative that policy makers and other users should be able to understand and act on any composite measure of performance. To that end, it recommends that, in parallel with technical improvements, WHO seeks vigorously to improve the capacity of users. Possible methods include:
- offering a transparent exposition of the data sources and methods used (limitations as well as advances);
 - presenting more disaggregate data as a means of 'drilling down' in order to understand better the components of a composite score, perhaps in the form of a balanced scorecard;
 - developing local analytic capacity.

- (viii) The natural starting point for performance assessment is a country's year-on-year change in attainment. SPRG views with some concern the likelihood that a country's composite index will change from year to year purely because of methodological changes. It therefore recommends that WHO should give careful consideration to how countries can be offered a useful time series of data which is not open to misinterpretation.

5. References

- Almeida, C. M., P. Braveman, M. R. Gold, C. L. Szwarcwald, J. M. Ribeiro, A. Miglionico, J. S. Millar, S. Porto, N. R. Costa, V. O. Rubio, M. Segall, B. Starfield, C. Travassos, A. Ugá, J. Valente, and F. Viacava (2001): Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357(9269): 1692-1697.
- Appleby, J. and A. Street (2001): Health system goals: life, death, and...football. *Journal of Health Services Research and Policy*, 6(4): 220-225.
- Coyne, J. S. and P. H. Hilsenrath (2002): The World Health Report 2000: Can health care systems be compared using a single measure of performance? *American Journal of Public Health*, 92(1): 30-33.
- Gakidou, E. E., C. J. L. Murray, and J. Frenk (2000): Measuring preferences on health system performance assessment. Global Programme on Evidence for Health Policy Discussion Papers, No. 20. Geneva, Switzerland: World Health Organization.
- Häkkinen, U. (2000): Assessment of the goal attainment and efficiency of health systems in the World Health Report 2000. In eds. U. Häkkinen and E. Ollila. *The World Health Report 2000: What does it tell us about health systems? Analyses by Finnish Experts*. Helsinki: National Research and Development Center for Welfare and Health (STAKES). [<http://www.stakes.fi/english/publicati/Publications.htm>]
- Hurst, J. and M. Jee-Hughes (2001): Performance measurement and performance management in OECD health systems. *Labour Market and Social Policy - Occasional Papers*, No. 47. Paris, France: Organization for Economic Co-operation and Development (OECD).
- Lancet (2001): Why rank countries by health performance? Editorial. *Lancet*, 357(9269): 1702-1703.
- Lauer, J., D. Evans, and C. J. L. Murray (2002): Measuring health system attainment: the impact of variability in the importance of social goals. Global Programme on Evidence for Health Policy, mimeo. Geneva, Switzerland: World Health Organization.
- McKee, M. (2001): Measuring the efficiency of health systems. *British Medical Journal*, 323(7308): 295-296.
- Murray, C. J. L., C. Mathers, J. Salomon, and A. Lopez (2001): Evidence for health policy: Dealing with missing data, uncertainty and ignorance. Global

Programme on Evidence for Health Policy, mimeo. Geneva, Switzerland: World Health Organization.

Murray, C. J. L., J. Lauer, A. Tandon, and J. Frenk (2000): Overall health system achievement for 191 countries. Global Programme on Evidence for Health Policy Discussion Papers, No. 28. Geneva, Switzerland: World Health Organization.

Navarro, V. (2000): Assessment of the World Health Report 2000. *Lancet*, 356(9241): 1598-1601.

Navarro, V. (2002): The World Health Report 2000: Can health care systems be compared using a single measure of performance? *American Journal of Public Health*, 92(1): 31-34.

Nord, E. (2002): Measures of goal attainment and performance: A brief, critical consumer guide. *Health Policy*, 59(3): 183-191.

Oswaldo Cruz Foundation, Brazilian Ministry of Health (2000): Report of the workshop "Health Systems Performance: The World Health Report 2000". 14-12-2000, Rio de Janeiro.

Rosén, M. (2001): Can the WHO Health Report improve the performance of health systems? *Scandinavian Journal of Public Health*, 29(1): 76-80.

Smith, P. (2002): Developing composite indicators for assessing overall system efficiency. In ed. P. Smith. *Measuring up: improving health systems performance in OECD countries*. Paris, France: OECD.

Ugá, A. D., C. M. Almeida, C. L. Szwarcwald, C. Travassos, F. Viacava, J. M. Ribeiro, N. R. Costa, P. M. Buss, and S. Porto (2001): Considerations on methodology used in the World Health Organization 2000 Report. *Cadernos de Saude Publica*, 17(3): 705-712.

Williams, A. (2000): Science or marketing at WHO? A Commentary on 'World Health 2000'. *Health Economics*, 10(2): 93-100.

Williams, A. (2001): Science or marketing at WHO? Rejoinder from Alan Williams. *Health Economics*, 10(4): 283-285.

World Health Organization (2000): *The World Health Report 2000. Health systems: improving performance*. Geneva, Switzerland: World Health Organization.

WHO Regional Office for Africa (2001): General Report on the Regional Consultative Meeting on Health Systems Performance Assessment: Final Report. 18-5-2001, Harare, Zimbabwe.

REPORT OF SPRG ON HSPA

WHO Regional Office for the Americas (2001): Regional consultation of the Americas on Health Systems Performance Assessment: Final Report. 8-5-2001, Washington, D.C.

WHO Regional Office for South-East Asia (2001): Report of the regional consultation and technical workshop on health systems performance assessment. 18-6-2001, New Delhi, India.

XII. DATA QUALITY AND DATA COLLECTION STRATEGIES

1. WHR 2000

The following comments pertain to data quality and data collection methods that were employed in WHR 2000. Data availability and data quality are critical issues for all the health systems performance measures in the report and supporting technical documentation and background papers. This brief discussion on data will be confined to broader data quality and availability issues, and will specifically comment on the World Health Survey (WHS). In-depth data issues that pertain to 'Responsiveness' and 'Fairness in Financing' are discussed in the respective sections.

2. Main commentaries and criticisms

Some of the strongest and most widespread criticisms of WHR 2000 related to data quality and availability (for example, Williams 2000; Oswaldo Cruz Foundation 2000; Ugá et al. 2000). The main strands of criticism were as follows:

- (i) that estimates were based on covariates in the absence of primary data on fairness-in-financial contribution, responsiveness, health inequalities, non-fatal health outcomes, death rates and life tables;
- (ii) that data were collected from key informants who may not be an appropriate source of information;
- (iii) that data requirements for HSPA are too onerous and resource intensive;
- (iv) that quality assurance was inadequate.

3. WHO responses and proposals

(i) Absence of primary data

For the most part WHO relied on datasets available within WHO, or on datasets consisting of national surveys, other surveys such as the DHS, and available household income and expenditure surveys. The only primary data collection efforts by WHO were a web-based survey to elicit information on the weighting of the different health goals, and a Key Informant Survey to obtain information on responsiveness in 33 countries. Acknowledging the limited scope of primary data collection for WHR 2000, new methods have been developed and surveys launched to improve data collection. This

includes the World Health Survey (WHS), which will be conducted in more than 70 countries. The methodology for WHS is based on the Multi-Country Survey Study 2000-2001 and is documented in Üstün et al. (2000).

(ii) Inadequacy of key informants approach

The key informant strategy is an inexpensive method of obtaining information on certain domains. WHO argue that for some domains, properly selected key informants may in fact provide more valid and less biased responses than the general population, owing to key informants' specific knowledge of these areas. In this approach the choice of key informants requires close attention so as to avoid possible biases.

In order to address this question empirically, WHO has collected data on responsiveness and health-system goals from key informants, selected through a snowball sampling technique from lists of health professionals and administrators. The set of questions asked were a subset of those in the questionnaire for the Multi-Country Survey Study, which canvassed the general population. As these two surveys were carried out in the same countries, WHO is able to compare the responses of the key informants and the general population, and address issues of systematic bias. Facility studies and exit interviews are also being planned by WHO to address issues concerning validity.

(iii) Data collection

To build consensus on data collection strategies and avoid duplication of efforts, WHO will collaborate with national agencies as well as with international organizations carrying out surveys such as DHS, LSMS, EURO Barometer and MICS. In addition WHO will provide technical support to countries or agencies wishing to include the WHO Survey instrument in whole or in part in their ongoing data collection strategies. It will build capacity in countries that request support for the introduction of quality assurance and analytical survey techniques.

(iv) Quality assurance

WHO is putting in place a range of quality assurance instruments. The World Health Survey is an important tool to support the quality assurance process. Within Member States a competitive bidding process has been put in place for the execution of WHS.

4. SPRG comments and recommendations

Data required to calculate all five components of the composite index for WHR 2000 were absent for most countries. Where data were available, the quality was not always of a high standard.

Documentation of methods and treatment of missing data

Most critiques of health systems performance assessment repeated the comment that the methodologies used, data sources, and assumptions made in the analyses of WHR 2000 were not adequately documented (Williams 2000; Almeida et al. 2001; McKee 2001).

SPRG recommends that, as a means of gaining transparency and confidence, WHO should make particular efforts to explain the treatment of missing data, and should discuss explicitly the assumptions and extrapolations used in the next round of HSPA.

Data not available where needed most

Data availability and data quality are even more of a challenge in countries whose health systems are not well established, where health information systems are rudimentary, or where health systems have collapsed for reasons of war or strife. It is usually the case that these environments are hard to reach, but these may be countries which have the greatest need for HSPA as a tool for change.

SPRG recommends that HSPA clearly needs to acknowledge this dilemma as a limitation, even though it is recognized that WHO cannot always overcome this difficulty.

SPRG recommends that WHO should make a deliberate effort for early implementation of the WHS in those countries and environments that have the least developed health-information systems.

Data collection

Wherever possible WHO should rely on existing data collection efforts within Member States and coordinate collection activity with the respective data agencies.

Collection of country statistics

The processes adopted in collecting and collating the data used in WHR 2000 are not adequately documented.

SPRG recommends that WHO helps to strengthen national data collection processes, including the government agencies that release official country statistics and data. This approach would immediately take care of potential disputes concerning the acceptability of the data used, but for validation purposes other sources should also be explored.

Data quality

Where data are available, their quality needs to be examined very carefully before any conclusions are drawn. Appropriate validation techniques should continue to be applied.

SPRG welcomes the WHO commitment to improved quality assurance methods. It also recommends that countries should participate in the interpretation and validation of the data to ensure that they are acceptable locally.

Key informants

SPRG considers the 'key informant' approach is susceptible to errors particularly for the HSPA exercise.

SPRG recommends that the 'key informant' approach should wherever possible be used alongside more objective sources of data. It also stresses that the choice of key informants needs careful consideration.

World Health Survey

WHS has been designed and developed on the basis of experience gained from the Multi-Country Survey Study 2000-2001, which was conducted in approximately 60 countries. WHS is designed on a modular basis with the intention of providing low-cost information that supplements data from national health-information systems to build up an evidence base for policy makers.

The commentaries all identify the need for reliable data and information as a basis for effective health-system monitoring. The African and European regional consultations emphasized the importance of WHS being closely aligned to national health-information systems (WHO Regional Office for Africa 2001; WHO Regional Office for Europe 2001). The SPRG view, based on discussion among colleagues and in country reports, is as follows.

- WHS is potentially very useful, and in broad terms SPRG welcomes its introduction.
- The WHS tool requires further refinement. The choice of modules needs to be reviewed for relevance.

XII. Data Quality and Data Collection Strategies

- SPRG recommends that appropriate links are established with other statistical offices and data collection initiatives.
- The survey teams should work closely with countries to ensure that integration with their health-information systems occurs in a meaningful way. While WHS will provide data for HSPA that are currently 'missing', it should not become another parallel system for collecting information that is used exclusively for the HSPA exercise. Rather, WHS should be seen as a mechanism to strengthen existing health-information systems.
- WHS should be self-sustaining, and should not compete for local resources.
- If WHS is to become an important instrument for strengthening national health-information systems, further consideration must be given to its sampling frame. Does the sampling frame enable conclusions to be drawn at sub-national level and comparisons made over time? Local needs must be taken into account in designing the WHS sampling frame.
- The WHS sampling frame should enable information to be obtained on vulnerable groups, such as refugees and itinerant and institutional populations. It is also important that the population covered by the Survey is representative of the population as a whole, including children (and especially girls).
- The issue of cross-population comparability is addressed in Section XIII.

5. References

- Almeida, C. M., P. Braveman, M. R. Gold, C. L. Szwarcwald, J. M. Ribeiro, A. Miglionico, J. S. Millar, S. Porto, N. R. Costa, V. O. Rubio, M. Segall, B. Starfield, C. Travassos, A. Ugá, J. Valente, and F. Viacava (2001): Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357(9269): 1692-1697.
- McKee, M. (2001): Measuring the efficiency of health systems. *British Medical Journal*, 323(7308): 295-296.
- Oswaldo Cruz Foundation, Brazilian Ministry of Health (2000): Report of the workshop "Health Systems Performance: The World Health Report 2000". 14-12-2000, Rio de Janeiro.
- Ugá, A. D., C. L. Szwarcwald, C. Almeida, C. Travassos, F. Viacava, J. M. Ribeiro, N. R. Costa, P. Buss, and S. Porto (2000): Views on the WHO 2000 Report. Rio de Janeiro: FIOCRUZ.
- Ustün, B., S. Chatterji, M. Villanueva, L. Bendib, C. Celik, R. Sadana, N. Valentine, C. Mathers, J. P. Ortiz, A. Tandon, J. Salomon, Y. Cao, X. W. Jun, and C. J. L. Murray (2000): WHO multicountry household survey study on health and responsiveness 2000-2001. Global Programme on Evidence for Health Policy Discussion Papers, No. 37. Geneva, Switzerland: World Health Organization.
- Williams, A. (2000): Science or marketing at WHO? A Commentary on 'World Health 2000'. *Health Economics*, 10(2): 93-100.
- Williams, A. (2001): Science or marketing at WHO? Rejoinder from Alan Williams. *Health Economics*, 10(4): 283-285.
- World Health Organization (2001): Background documentation: Scientific Peer Review Group meeting. Geneva, Switzerland.
- WHO Regional Office for Africa (2001): General Report on the Regional Consultative Meeting on Health Systems Performance Assessment: Final Report. 18-5-2001, Harare, Zimbabwe.
- WHO Regional Office for Europe (2001): Report of the Regional Consultation on Health Systems Performance Assessment. 3-9-2001, Copenhagen.
- WHO Regional Office for South-East Asia (2001): Report of the regional consultation and technical workshop on health systems performance assessment. 18-6-2001, New Delhi, India.

XIII. CROSS-POPULATION COMPARABILITY

1. WHR 2000

In making the estimates for WHR 2000, corrections were made for major known biases in available measurements to improve cross-population comparability – for example, for under-reporting of mortality data in vital registration systems. The concept of internal consistency was used as a tool to improve the validity of epidemiological assessments.

2. Main commentaries and criticisms

Data criticisms of WHR 2000 were rather severe but this section deals only with the question of cross-country comparability. There has been little public debate and discussion on this issue beyond recognizing it as a problem with self-report data.

3. WHO responses and proposals

In examining self-assessed morbidity from survey data across the states of India, Murray and Chen (1992) reported the following findings: Kerala has the highest self-reported morbidity, and Bihar the lowest, across the Indian states. On the other hand, an objective measure of health – such as mortality – reveals that Kerala has a much higher life expectancy than Bihar. Next, a comparison between the US and Kerala shows that self-assessed morbidity in the US is much greater than in Kerala, despite life expectancy in the US being higher than that in Kerala.

What is going on? Are there features of the environment – educational, medical (e.g., frequency of exposure to the health system), income, etc. – that can explain these apparently inconsistent findings? Amartya Sen (1992) in an article in *Philosophy and Public Affairs* tried to understand these results in terms of what he called 'positional objectivity': the 'position' of the individual (in terms of education, income, etc.) matters in the response that is given – but all individuals in the same position will give the same response – hence 'positional objectivity'. In a more recent editorial in the *British Medical Journal*, he again emphasizes the fact that self-reported morbidity data have limitations that can make its use extremely misleading for policy purposes (Sen 2002).

WHO is seeking to make the responses of individuals comparable (whether they live in different states of India or in the US) by *correcting* for the 'positions' of the individuals in the different states of India and the US. This is obviously a very important exercise in obtaining health-status information from survey data that is comparable across countries. Moreover, self-reported data on health are still by far the most common source of such information around the world.

As a response to the paucity of representative population-based information on two key variables in the HSPA exercise in WHR 2000, the WHO launched the Multi-Country Survey Study on Health and Responsiveness. For the purposes of HSPA, these survey data are utilized to construct measures of: (i) health-adjusted life expectancy (HALE), and (ii) the level of responsiveness of the health system in a country. For example, the measurement of HALE includes estimates of non-fatal health that are, in part, derived from survey data on the different domains of health (e.g., mobility, cognition, affect, etc.). Similarly, the level of responsiveness of a country's health system is also based on such survey data. Respondents are asked to evaluate their experiences relating to different domains of responsiveness of the health system (e.g., autonomy, dignity, prompt attention, etc.).

There are two characteristics of these survey data that lead to the problem of cross-population comparability. First, the information on the domains is obtained on the basis of self-reporting. Respondents are asked to evaluate their own experience (or perception) with respect to various domains of health and of health-system responsiveness. Secondly, these self-report responses are categorical and ranked ordinally.

One example from the WHO Multi-Country Survey Study for the health domain of mobility illustrates the characteristics of the data. The main self-report question asks respondents how much difficulty they have had in moving around in the past 30 days. Respondents are asked to characterize their mobility using a 5-category ordinal response scale ranging from 1 to 5, where 1 is "Extreme/Cannot do", 2 is "Severe difficulty", 3 is "Moderate difficulty", 4 is "Mild difficulty", and 5 is "No difficulty".

This is where the issue of cross-population comparability arises. The problem with using these self-report data from the domains of health and responsiveness is that the responses are not comparable across countries, or even across different socio-demographic groups within countries. As Figure 1 illustrates, the categorical responses can be conceptualized as a mapping from the true level of the domain (here the line labelled "latent mobility scale") to the categorical responses for three different populations A, B, and C. As the figure shows, someone answering "No difficulty" in population A maps to a different interval on the true scale as someone answering "No difficulty" in populations B and C. Obversely, the same level of true mobility could be self-reported by a person in population A as representing "no difficulty", by a person in population B as representing "mild" difficulty, and

by a person in population C as representing “moderate” difficulty. The reasons could be due to differing norms, expectations, and experiences of respondents from different populations.

This problem has been previously identified in the psychometrics literature on ability (IQ) testing and, more generally, in educational testing through standardized tests (e.g., GRE, SAT, GMAT, etc.). Certain groups, conditional on ability or knowledge, systematically do better on certain types of question than other groups. This problem is known as “differential item functioning” in the educational testing and psychometrics literature (Holland and Wainer 1993). For instance, in the item response theory literature, the partial credit model (which is akin to the ordered probit model) specifies the probability of responding in one of two ordered (adjacent) categories as an increasing function of a respondent’s ability and a decreasing functioning of the category difficulty. For the same level of ability, the difficulties may be systematically different for different population groups, which will lead to a bias in measured ability. Although this problem is similar to the problem of cut-point shifts in measuring health or health-system responsiveness, the solution methods are somewhat different.

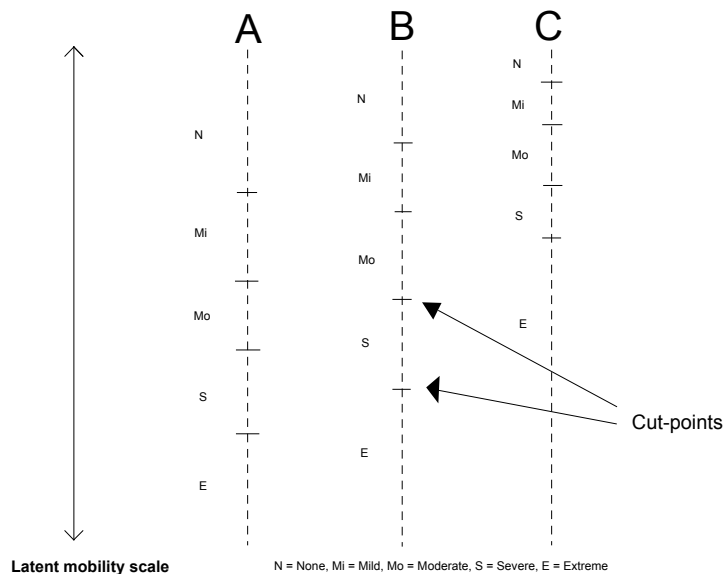


Figure 1

There are basically two strategies that WHO has developed to adjust survey responses for systematic differences in people’s attitudes. Both strategies involve the use of a statistical model – the hierarchical ordered probit (HOPIT) model. The first strategy is to use the HOPIT model with ‘vignettes’. The second strategy is to use the HOPIT model with measured tests. These are described in turn.

A vignette is a description of a level of ability on a given domain that respondents are asked to evaluate with respect to the same question and the on the same categorical response scale as the main self-report question. A vignette depicts a fixed level of ability on a given domain, so that for that vignette, differences in responses across countries or socio-demographic groups may be attributed to differences in cut-points for the response categories. The response category cut-points are estimated by use of the HOPIT model through a maximum likelihood procedure. These cut-point estimates are used to calibrate the respondent's own self-report in order to make it cross-population comparable. If, for example, respondents from a certain population group systematically give higher categorical responses to the vignettes than respondents from another group, this will show up as a lower cut-point for the first group in the HOPIT estimation.

A second strategy is to calibrate self-report responses using measured tests (instead of vignettes) in conjunction with the HOPIT model. Measured tests are tests of the level of ability of the underlying latent variable for a domain of health. Examples include the posturo-locomotion-manual (PLM) test for mobility, and the Snellen eye chart exam for the domain of vision. Such measured tests are used to estimate cut-point differences across population groups for calibration of self-report responses that are cross-population comparable. The set-up of the HOPIT model with measured tests is quite straightforward. The model assumes that the measured test is correlated with the underlying latent variable for a domain, and the cut-points for a particular categorical response are allowed to differ by population group.

For a variety of reasons including those related to measurement error, it appears that vignettes are a superior mechanism than measured tests for the calibration of self-report responses. Hence, current WHO estimates of outcome measures of health and responsiveness are based primarily on the use of vignettes as a calibration strategy.

4. SPRG comments and recommendations

SPRG welcomed WHO's work in this area and recognized the importance of ensuring that the data used in the HSPA exercise are comparable across populations.

- (i) The HOPIT methodology depends crucially on the assumption that the categorical responses derive from a single dimension (or attribute), which can be ordered on a unilinear scale. SPRG noted the responses should not be based on mappings by individuals that involve comparisons in two or more dimensions (of planar or higher-dimensional regions corresponding to the five categories). Application

of this methodology requires that the domain of each self-report question is narrowly and unambiguously specified.

- (ii) A promising avenue for future research would be to develop statistical methods that combine the information from both vignettes and measured tests in a joint estimation procedure. These methods, akin to the multiple-indicator multiple-cause models in the statistical literature, have the advantage that they take full account of all available information on a given individual, in this case the multiple sources being the individual's self-report (calibrated using vignettes) as well as his/her measured test. These types of methods can also allow for different statistical errors in information that is self-reported and information that is obtained from measured tests.
- (iii) The HOPIT model depends critically on the cross-cultural reliability and consistency of the vignettes – e.g., translation problems or errors do not change the meaning of a question so that a different latent variable is being measured. SPRG recommends that the vignettes be tested further in different settings, including through back translation.
- (iv) It may be possible to explore some of the problems related to (iii) above through a random coefficients version of the HOPIT model. Unlike the current version of the model, a random coefficients model allows the latent variable associated with each vignette to have its own variance (rather than the variance being the same for all vignettes). This method allows one to take account of the possibility that some vignettes may be inherently 'noisier' than others. This may be of particular relevance for vignettes referring to the middle range(s) of a domain, i.e., for vignettes that are not at either extreme of a domain.
- (v) SPRG noted that the HOPIT model not only addressed the problem of cross-population comparability, but also converted the discrete (categorical) information on each domain of health and responsiveness into a *continuous variable*. For each individual the aggregation of these variables across the appropriate domains generates the continuous distribution from which the mean level of, and inequality in, health (or responsiveness) is estimated. Hence, the HOPIT model yields much more than cross-population comparability: it forms the basis for estimating four of the five intrinsic indicators used in HSPA.
- (vi) SPRG members made several technical comments on the HOPIT methodology. Some of these are noted below.
 - (a) The estimates of the cut-points for a population group (e.g., country) will depend on the universe of groups included in the cut-point estimation. For example, suppose the cut-points for group A are estimated from data for groups A and B. Now, data on group C become available and the cut-points for A are re-estimated from data

for all three groups A, B, and C. In general, the cut-points (and other parameter estimates) for group A will change. This could make the relative ranking between, say, groups A and B depend on the precise other groups included in the estimation (especially when considering the aggregates across domains). Hence, caution will need to be exercised in making judgements about the relative ranking between countries, which could be universe-dependent.

(b) The assumption made in the HOPIT model is that the latent variable (e.g., mobility) is unbounded (as the normal distribution is used for the error term). SPRG recommends that the WHO Secretariat check the robustness of their results to restricting the latent variable to a finite interval (e.g., through the assumption of a truncated normal distribution for the error term), as this would seem a more realistic assumption for the domains considered.

(c) SPRG members commented that it would be valuable to estimate non-linear functional forms for the latent variable equation (e.g., health production function), which might also to some extent address the problem noted in (b). A log-linear form for the health production function seems more realistic as it allows for diminishing returns to the factors that determine health (e.g., age, education, etc.), which may be more reasonable than assuming constant returns to each factor. In any case, it would be valuable to check the sensitivity of the present HOPIT results to the assumptions made about the functional form.

5. References

Holland, P. W. and H. Wainer (1993): *Differential Item Functioning*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Murray, C. J. L. and L. C. Chen (1992): Understanding morbidity change. *Population and Development Review*, 18(3): 481-503.

Sen, A. K. (2002): Health: Perception versus observation. *British Medical Journal*, 324: 860-861.

Sen, A. K. (1992): Positional objectivity. *Philosophy and Public Affairs*, 22: 126-145.

XIV. EFFICIENCY

1. WHR 2000

This section provides a commentary on the methodology used to measure health-system efficiency in the WHR 2000 statistical analysis, and on the changes proposed for the future. The methods used in WHR 2000 are outlined in World Health Organization (2000), pages 40-44, and the results given in Annex Table 10. Further details of the methods used are provided in Evans et al. (2000) and Tandon et al. (2000). The methodology was subsequently discussed at a WHO meeting of experts (World Health Organization 2001), and the analysis of efficiency with respect to healthy life expectancy has been reported in Evans et al. (2001).

To measure efficiency WHR 2000 used a frontier production function approach, an established technique employed to assess the efficiency of agricultural or industrial production, but which has been extended to the areas of education, local government, and health. This technique estimates the relationship between output and inputs to production, and the highest possible output that could have been produced for each combination of inputs. The ratio of the observed output to the maximum that could have been produced is defined as the efficiency score.

WHO modified this technique to allow for the fact that health outcomes in the absence of a functioning system would still be positive, not zero. So a minimum output level corresponding to the absence of health-system inputs was also estimated, using the relationship observed between literacy and health outcomes in the early 1900s. The inputs used to estimate efficiency in WHR 2000 were health expenditure per capita and average years of schooling of the adult population. Efficiency was estimated for the overall attainment index as well as for health attainment (HALE) separately.

The term "efficiency" is used throughout to denote the level of attainment secured by the health system in relation to spending and environmental inputs (external influences on attainment). In WHR 2000 this concept was referred to as "performance", but on the basis of the regional consultations WHO has decided to denote it as "efficiency".

2. Main commentaries and criticisms

The approach to efficiency measurement used in WHR 2000 is based on the parametric frontier estimation methods traditionally used in productivity

analysis. These are analogous to usual statistical regression analyses, except that the 'error' term for any observation may be decomposed into two elements – the conventional two-sided random error, and a one-sided error attributable to inefficiency. Such productivity models have reached an advanced stage of econometric development, and have been applied in a number of different areas. The expert group assembled by WHO included some of the leading exponents of productivity modelling. It broadly endorsed the statistical approach used in WHR 2000, but it should be noted that there are those who contest the entire edifice on which modern productivity modelling is based (Newhouse 1994; Stone forthcoming).

Criticisms of the WHO methods can be considered under four headings: philosophical concerns, the theoretical production model, measurement issues, and estimation. As well as the published commentaries listed in the references below (e.g., Häkkinen 2000; Jamison and Sandbu 2001; McKee 2001; Navarro 2002; Oswaldo Cruz Foundation 2000; Gravelle et al. 2002; Grignon 2001; Richardson et al. 2002; Hollingsworth and Wildman 2002; etc.), we have also seen a number of as yet unpublished papers that we are unable to cite. We are grateful to these authors for the privileged access to their material. The concerns are listed without an attempt to judge their validity. SPRG comments and recommendations follow under heading 4 below.

Philosophical concerns

- The econometric methods used to estimate efficiency are both complex and relatively new. This makes understanding and interpretation difficult, especially for nations seeking to improve their health-system performance (Almeida et al. 2001).
- The use of the concept of efficiency may send a confused message when set alongside the objective of improving health outcomes. A country might have low absolute levels of health attainment but still be deemed technically efficient because it spends very little on its health system. The concept of efficiency makes no judgement about how much should be spent on health, but health outcomes can evidently be improved by higher expenditure.
- The determinants of health-system performance are too complex to be reducible to a tractable statistical model, particularly in view of the poor quality of the data, the relatively small number of observations, and the lack of reliable time-series information.
- Parametric statistical models traditionally focus on estimating the relationship between a stimulus (inputs) and a response (in this case, attainment) but not on the residual for an individual observation. In contrast, productivity models concentrate on these residuals, and therefore require much greater attention to be placed on model specification.

- In particular, it can be argued that – in an application as complex as the WHO endeavour – it is inevitable that there is significant measurement error and that the model specification is incomplete. In these circumstances, little confidence can attach to the estimated measure of inefficiency.
- The uncertainty analysis used by WHO is incomplete, as it does not fully consider modelling errors that are potentially important sources of uncertainty (see also Section XI).
- Despite the progress made, there are numerous unresolved issues surrounding the methodology of productivity analysis. It may be premature to base definitive rankings of health systems on such developmental methodology.

The model of production

- The technical consultation on 'Measurement of Efficiency of Health Systems' seemed to be comfortable with the use of a single production function (on the grounds that all countries have access to the same medical technologies). However, several commentators have argued that the health production function may not be identical between nations, suggesting that there is disagreement about whether the use of a single model is appropriate (Richardson et al. 2002; Häkkinen 2000; Nord 2002).
- More generally, there is no consensus that the WHO approach uses an appropriate theoretical model of the production process it seeks to capture (Pedersen 2002; Grignon 2001). For example, many of the outcome indicators in WHR 2000 are influenced strongly by factors other than the health system (e.g., war or diet), and these are inadequately captured in the WHO model of production. The treatment of income has generated particular debate. It is also argued that some of the outcome indicators – e.g., health inequality – are affected not just by the average level of inputs (e.g., education, income) but also by the distribution of inputs (inequalities in education, income) (Ammar and Awar 2001).
- Although the proposed work on functions of the health system may help in the future, the methods used in WHR 2000 do not adequately model the 'reasons' why a given level of efficiency is observed (Grignon 2001; Pedersen 2002).
- The chosen model does not recognize the important time lags that exist in producing health outcomes (Grignon 2001).
- The need to calculate a "minimum" level of health attainment in the absence of a health system is contested (Gravelle et al. 2002; Häkkinen 2000).

Measurement issues

- The description of the treatment of missing data is inadequate, as in the HSPA exercise as a whole (see Section XII). Estimates of missing data

- will be subject to considerable errors-in-variables, and hence will cause biases in parameter estimates and possibly in rankings (Almeida et al. 2001; Häkkinen 2000; Pedersen 2002).
- The components of the efficiency model refer to different definitions of the health system – for example, the output measure refers to a very broad definition of the health system, whilst the input (expenditure) measure relates predominantly to expenditure on health care (Nord 2002).
 - The composite measure of output is highly contested and embraces numerous assumptions and value judgements (see Section XI), which have consequential implications for the efficiency measure.
 - Relative prices of inputs differ between nations, and estimates of total expenditure do not reflect the cost advantages in producing different outputs.
 - The measures of cost rely on PPP-adjusted estimates of expenditure, which are subject to error (in the absence of health-specific PPP factors), causing bias in parameter estimates and possibly in country ranks (Grignon 2001).
 - Years of education is an inadequate proxy for external influences on health-system performance (Williams 2001; Jamison and Sandbu 2001; Häkkinen 2000; Grignon 2001).
 - The methodology and data used to measure the “minimum” are contested (Williams 2001; Pedersen 2002).

Econometric methodology

- The use of the fixed-effects panel data estimator is inappropriate, given the very low degree of variation from one year to the next in most observations (Gravelle et al. 2002).
- The models used presume a fixed level of efficiency across the entire four-year period examined, which may be unrealistic (Pedersen 2002; Gravelle et al. 2002).
- The methods do not adequately treat the important contribution of income to the production of health and therefore to health-system performance. The role of income needs to be properly modelled even if estimation turns out to be econometrically inconvenient (because income is highly correlated with both inputs and outputs) (Pedersen 2002).
- Formal model-selection techniques should be employed in choosing the preferred functional form for the model.
- More details are required on whether the chosen model passes the usual model misspecification statistical tests.
- There is evidence of a structural difference between developed and less-developed countries, implying the need for separate modelling (Richardson et al. 2002).
- Equally plausible alternative statistical model specifications can give rise to significantly different results (Gravelle et al. 2002; Hollingsworth and

Wildman 2002; Richardson et al. 2002; Grignon 2001; Jamison and Sandbu 2001).

3. WHO responses and proposals

The following detailed issues are highlighted in Section VI of the WHO Summary Document and in discussion documents prepared for SPRG, which include WHO proposals for further development of this work. WHO proposes to continue developing the concept of efficiency on the grounds that health resources are scarce in all Member States. The Secretariat believes it is important to determine if those resources contribute to the greatest extent possible to the outcomes that people value. This is an important complement to the goal of finding additional resources for health.

(i) Timing

There are lags between the timing of health-system inputs and health outcomes. In WHR 2000 the assumption was made that current expenditures are highly correlated with past expenditures, but it would be preferable to use a time series of expenditures in explaining health outcomes and measuring efficiency. Data limitations prevented this in WHR 2000.

(ii) The minimum

In the absence of a health system population health (e.g., life expectancy at birth) would still be positive, so WHO argues that it is important to identify the minimum level. The minimum for WHR 2000 was estimated from limited data around 1900 when the modern health system did not exist. Only literacy was found to be correlated with health outcomes, but it would be useful to determine if there are other ways of defining the minimum.

(iii) Difficulty

Variations in the difficulty of translating inputs into outcomes were not fully captured in the production function in WHR 2000. Some, however, were subsequently analysed in the second-stage analysis.

(iv) Determinants of output

There is ongoing debate about the correct specification of the production function, but the Secretariat argues that it is critical to separate clearly the inputs to production from the factors that influence the efficiency of the production process.

(v) Determinants of efficiency

The technical consultation on this topic had suggested that the determinants of efficiency were better estimated at the same time as the estimation of efficiency rather than at a second stage, which was the approach adopted by WHO.

Because of the complexity of the issues surrounding efficiency, WHO has proposed some new analysis including: (a) estimating the traditional production function as a one-step process for efficiency simultaneously with

the possible determinants; (b) a random-coefficients econometric specification of the production model. In terms of a new approach, WHO has proposed that the unobserved efficiency variable could be inferred from a multiple indicator model, which would use both the existing specification and additional models based on measures of process, such as coverage.

In light of this debate WHO proposes the following.

- (i) The questions of timing and how best to estimate the minimum are complex and the opinions of relevant outside experts will be sought. At the same time, the multiple indicator model approach is promising and should be developed further. This requires that the proposed work on coverage, discussed in Section IV of the SPRG report, should continue (with the World Health Survey providing the relevant information).
- (ii) To address the question of timing, following discussion with SPRG two suggestions were made. The first was to estimate current HALE as a function of the series of past expenditures, or to use HALE at some time in the recent past, say five years ago, as a controlling variable. (The latter method has the drawback of lagged dependent variable models, while the former requires developing a historical time series of health expenditures.) The second proposal was to pursue the question of incidence-HALE – the HALE that is determined by this year's activities (Section V.A in the Summary Document). 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.

4. SPRG comments and recommendations

SPRG considers that there are strong arguments in favour of seeking to measure health-system efficiency. Consideration of efficiency should – in principle – permit valid comparison of systems operating with different health expenditures and in different external environments. It could therefore make a vital contribution to HSPA. The WHO initiative has launched interesting technical debates and a research agenda that has the potential to advance rapidly our state of knowledge of health-system performance. It has also stimulated the search for improved conceptual models and data sources, and has made some innovative technical contributions to productivity analysis.

However, there are some important objections in principle to the method used by WHO, the most important of which are: (a) that the health system is too complex to be captured by these simple statistical models; (b) the data available are currently inadequate to support such an endeavour; and (c) the

analysis is too demanding technically to be helpful to policy makers and other government officials.

In addition, it is possible to invoke numerous practical objections to the methods that have been applied, many of which are summarized above. Some of the most important are: (a) the treatment of missing data (applies to all of HSPA); (b) the treatment of influences on outcomes other than the health system; (c) the inadequate treatment of time lags; (d) the method of implementing some of the econometric techniques used; (e) the handling of uncertainty and sensitivity analysis.

We also believe that there should be complete 'transparency' of the research process relating to efficiency. As in much econometric work, the findings in WHR 2000 are the result of numerous technical judgements, and are not just the consequence of ineluctable scientific logic. Examples include the nature of the model of production, the concept of the "minimum", the treatment of missing data, and a series of econometric choices. We recognize that there may not always be consensus regarding the correct technical approach. However, it is in our view imperative that all technical judgements are capable of being understood, scrutinized and challenged by external observers. This requires preparation of a technical audit trail, publication of all methods used, and ready availability of data.

Recommendations

As the debate on WHR 2000 has demonstrated, any analysis of efficiency must be considered work-in-progress rather than a definitive judgement on health systems. On balance, we feel there is a case for continuing work in this area. However, we recommend that any continued WHO work on health-system efficiency should be presented as an ongoing research programme rather than a definitive judgement on health systems, and that progress should be reviewed at regular intervals.

The practice of publishing a league table of nations based on efficiency estimates has been highly contentious, but in the view of SPRG the decision to continue publishing league tables is a strategic and policy decision for WHO rather than a scientific one. Given the large number of technical problems that have still to be resolved, we recommend that this work should be developed further, and that any tables produced should be recognized as work-in-progress.

There are numerous possibilities for improving the data sources on which the efficiency rankings are based. These include improvement in the measurement metrics and the treatment of missing data (considered elsewhere in this report), and where possible the use of sub-national data sources. We recommend that WHO should make strenuous efforts to improve the quality and extent of data used in efficiency analysis (indeed in

all of HSPA), and to adopt a transparent and careful approach to the treatment of missing data.

Particular concern has been expressed in the literature at the comparison of all health systems within a single model of production. It is possible that systems in different environmental circumstances are confronted with different production possibilities. We recommend that WHO should carefully explore the implications of incorporating environmental factors into the analysis, or developing separate models for different types of health system.

A particular conceptual weakness of methods to date has been the treatment of time. Measures of health outcome reflect years of health-system endeavour, while measures of expenditure refer to the current period. Furthermore, health-outcome measures are likely to be affected by factors other than the health system. These weaknesses suggest that contemporary measures of future (predicted) outcomes, e.g. certain process measures, may be more satisfactory measures of system performance than health-outcome measures. For this reason, we recommend that WHO should explore the scope for incorporating coverage and other measures of process into the model of efficiency.

The econometric analyses presented in WHR 2000 and subsequent variants exhibit some scientific weaknesses. We recommend that WHO engages in an ongoing consultative process with relevant experts to address the technical issues raised by outside commentators.

The treatment of uncertainty in WHR 2000 needs to be expanded as does the sensitivity analysis that was presented. We recommend that the method of modelling and presenting uncertainty should be reformulated to include a much broader scope of alternative models and assumptions.

The issues surrounding the measurement of efficiency are undoubtedly complex and require extensive data. Because of this complexity, we feel that in this area – perhaps more than in others – the input of a wide range of experts from different backgrounds is desirable. We recommend that WHO should actively consult and engage outside experts in the further development of this area, and that its analyses should be fully documented to maintain transparency.

5. References

- Almeida, C. M., P. Braveman, M. R. Gold, C. L. Szwarcwald, J. M. Ribeiro, A. Miglionico, J. S. Millar, S. Porto, N. R. Costa, V. O. Rubio, M. Segall, B. Starfield, C. Travassos, A. Ugá, J. Valente, and F. Viacava (2001): Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357(9269): 1692-1697.
- Ammar, W. and M. Awar (2001): What does the World Health Report bring to Lebanon? *Lebanese Medical Journal*, 49(3).
- Appleby, J. and A. Street (2001): Health system goals: life, death, and...football. *Journal of Health Services Research and Policy*, 6(4): 220-225.
- Coyne, J. S. and P. H. Hilsenrath (2002): The World Health Report 2000: Can health care systems be compared using a single measure of performance? *American Journal of Public Health*, 92(1): 30-33.
- Evans, D., A. Tandon, C. J. L. Murray, and J. Lauer (2000): The comparative efficiency of health systems in producing health. Global Programme on Evidence for Health Policy Discussion Papers, No. 29. Geneva, Switzerland: World Health Organization.
- Evans, D., A. Tandon, C. J. L. Murray, and J. Lauer (2001): Comparative efficiency of national health systems: cross national econometric analysis. *British Medical Journal*, 323: 307-310.
- Gravelle, H., R. Jacobs, A. M. Jones, and A. Street (2002): Comparing the efficiency of national health systems: econometric analysis should be handled with care. York: University of York. mimeo.
- Grignon, M. (2001): A southern European paradox: could reducing education budgets spur improvement in physicians performance? A propos World Health Report 2000. Paris: CREDES. mimeo.
- Häkkinen, U. (2000): Assessment of the goal attainment and efficiency of health systems in the World Health Report 2000. In eds. U. Häkkinen and E. Ollila. *The World Health Report 2000: What does it tell us about health systems? Analyses by Finnish Experts*. Helsinki: National Research and Development Center for Welfare and Health (STAKES).
<http://www.stakes.fi/english/publicati/Publications.htm>
- Hollingsworth, B. and J. Wildman (2002): The efficiency of health production: Re-estimating the WHO panel data using parametric and non-parametric approaches to provide additional information. Melbourne: Monash University. mimeo.

Jamison, D. T. and M. E. Sandbu (2001): Global health: WHO ranking of health system performance. *Science*, 293(5535): 1595-1596.

McKee, M. (2001): Measuring the efficiency of health systems. *British Medical Journal*, 323(7308): 295-296.

Navarro, V. (2000): Assessment of the World Health Report 2000. *Lancet*, 356(9241): 1598-1601.

Navarro, V. (2002): The World Health Report 2000: Can health care systems be compared using a single measure of performance? *American Journal of Public Health*, 92(1): 31-34.

Newhouse, M. (1994): Frontier estimation: how useful a tool for health economics. *Journal of Health Economics*, 13: 317-322.

Nord, E. (2002): Measures of goal attainment and performance: A brief, critical consumer guide. *Health Policy*, 59(3): 183-191.

Oswaldo Cruz Foundation, Brazilian Ministry of Health (2000): Report of the workshop "Health Systems Performance: The World Health Report 2000". 14-12-2000, Rio de Janeiro.

Pedersen, K. M. (2002): The World Health Report 2000: Dialog of the deaf? *Health Economics*, 11: 93-101.

Richardson, J., I. Robertson, and J. Wildman (2002): A Critique of the World Health Organisation's Evaluation of Health System Performance. Melbourne: Monash University. mimeo.

Rosén, M. (2001): Can the WHO Health Report improve the performance of health systems? *Scandinavian Journal of Public Health*, 29(1): 76-80.

Stone, M. (forthcoming): How not to measure the efficiency of public services (and how one might). *Journal of the Royal Statistical Society, Series A*.

Tandon, A., C. J. L. Murray, J. Lauer, and D. Evans (2000): Measuring overall health system achievement for 191 countries. Global Programme on Evidence for Health Policy Discussion Papers, No. 30. Geneva, Switzerland: World Health Organization.

Williams, A. (2000): Science or marketing at WHO? A Commentary on 'World Health 2000'. *Health Economics*, 10(2): 93-100.

World Health Organization (2000): *The World Health Report 2000. Health systems: improving performance*. Geneva, Switzerland: World Health Organization.

World Health Organization (2001): Report on WHO meeting of experts on the measurement of efficiency of health systems. 1-8-2001, New Orleans.

XV. ENHANCING POLICY RELEVANCE

1. WHR 2000

The results of the performance assessment exercise in WHR 2000 were presented in the Statistical Annex of the Report as:
attainment on the five intrinsic goals separately;
attainment on the composite index;
efficiency in terms of average health level and in terms of the composite index.

All scores were presented in rank order, with uncertainty intervals around the scores and ranks. In the overview and the first two chapters of the Report there was discussion of the potential policy uses of quantitative analysis of the health-system goals. The four chapters on functions reviewed current evidence on the relation between outcomes and provision, resource generation, stewardship, and different ways of financing. The Report did not provide country-specific interpretation of this analysis but drew general conclusions about the type of strategies which will enhance performance.

2. Main commentaries and criticisms

The policy uses of ranking

There was mixed feeling at the regional consultations about the value of publishing overall attainment and efficiency scores and the accompanying rankings. This topic is discussed elsewhere (see Sections XI and XIV). In relation to rankings, some participants in regional consultations argued that the overall attainment and efficiency estimates should not be reported country-by-country. An alternative suggestion was to group countries by level of attainment, e.g. high, medium and low. However, other participants saw the value of ranking as a means of focusing the attention of policy makers on the health system and its performance.

Rosén (2001) welcomed the “underlying idea of generating a discussion on how well health systems function in different countries by openly reporting comparative statistics”. Navarro (2001) takes a similar stance. Appleby and Street (2001) comment on how information, and ranking in particular, may be used in different ways by policy makers and the public. In the Summary Document prepared for SPRG, WHO states that: “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.”

Multiple goals

The fact that the WHO framework explicitly recognizes there are multiple goals for a health system has been welcomed as being useful for policy purposes (Walt and Mills 2001; McKee 2001). Appleby and Street (2001) particularly note that it is useful in thinking about trade-offs between goals.

Procedural concerns

Other comments were of a procedural nature. For example, government officials 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 are 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 either ignored or misunderstood by the media.

The need to increase confidence in and ownership of results

The perceived policy relevance of the results was partly affected by concerns about the data and the methods used (see other Sections). In addition, it was noted that confidence and ownership of results would be enhanced by more national involvement in method development, and improved capacity to apply the complex methods and tools and to interpret the results.

Summary measures of outcomes are not sufficient for policy development purposes

The measures provide information on ‘how well’ a health system is performing, but not on ‘why’ it is performing as it is. Appleby and Street (2001) observe that if countries are to respond positively to HSPA, this involves finding variables that both explain performance and are open to policy manipulation. It is argued that additional information on determinants and on intermediate goals related to health-system functions, such as access, is essential for policy development (Almeida et al. 2001; Braveman et al. 2001; Makinen et al. 2000; Nord 2002; Van der Stuyft and Unger 2000; WHO Regional Office for the Americas 2001; WHO Regional Office for Africa 2001; WHO Regional Office for the Eastern Mediterranean 2001; WHO Regional Office for Europe 2001; WHO Regional Office for South-East Asia 2001).

Sub-national or sub-system analyses may also be needed for policy analysis and development (Wibulpolprasert and Tangcharoensathien 2001).

The need for an explicit strategy to link assessment to policy dialogue and system development.

Since the publication of WHR 2000, many have argued that the links between the measurement of performance and the development of policy requires strengthening.

A number of commentators have observed that availability of relevant information does not necessarily lead to its 'use' (Kvale 2000).

3. WHO responses and proposals

Since the publication of WHR 2000, a number of countries have 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. Participants in the regional consultations also emphasized that the links between the measurement of performance and the development of policy required strengthening. To meet the country requests and suggestions of the regional consultations, the Director-General decided to group efforts under the rubric of 'Enhancing Health Systems Performance Initiative'(EHSPI). Around 30 Member States expressed interest. WHO is currently working with 20 countries from different income ranges and WHO Regions. Reasons for engaging include:

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 policy, particularly related to the four functions;
- the search for greater contact with and opportunities for learning about health systems in other countries.

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;
- 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:

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

Strategies to meet these objectives are discussed under the following four headings.

(i) Describing and understanding health-system performance

Outcomes

Initial work 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 required, and for refining the indicators. For example, some countries have been interested in testing what mode of survey is the most cost-effective in obtaining the desired information, so more than one modality has been tested. Others have provided feedback which has helped to modify the indicators of the intrinsic goals or the questionnaire of the World Health Survey.

Inputs

To date, most attention has focused on improving estimates of health expenditures in countries lacking national health accounts. There are several regional initiatives supporting National Health Accounts (NHA) construction, and EHSPI has facilitated their support. A *Producer's Guide to National Health Accounts for Low and Middle Income Countries*,¹ jointly authored by WHO, the World Bank and USAID, will soon be published in English, French and Spanish. This interaction with NHA networks seems the most appropriate way to build evidence in this respect.

Functions

A number of countries have requested help to measure the performance of the four basic functions in their country settings. To this end, the major emphasis has been on defining an indicator of service provision that is more useful for policy than geographical access. A number of countries are testing the new WHO tool to measure effective coverage.

(ii) Implementation: linking evidence to policy

A number of participant countries have held national seminars to introduce a wider range of decision makers and researchers to the performance

¹ These will appear in the NHA *Producer's Guide* currently under preparation. It is co-funded by the World Bank, WHO and USAID, and is being jointly prepared by those agencies and a team from the Harvard School of Public Health.

assessment approach, and to discuss the policy implications of findings from the baseline analysis. In addition, WHO is providing direct policy support to a small number of countries, incorporating the new information being generated from these efforts.

(iii) Sub-national performance assessment

Some have suggested that the assessment framework 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. There will be an international meeting in 2002 to discuss the practical and methodological challenges in adapting the framework.

(iv) Building capacity in health-system performance assessment and analysis

For health systems performance assessment to be sustainable at the country level, capacity for both the diagnostic and implementation phases must be built. A variety of strategies have been piloted – ranging from straightforward briefings on using the methods, to technical support to analysts in-country or at WHO, to formal training workshops. There have been international workshops in South Africa and Indonesia (in English), in China, and one for French speakers is scheduled in Africa in 2002.

WHO proposals for increasing policy relevance

(i) Increasing the knowledge base on health systems

Helping Member States monitor health-system performance

WHO will support the generation of better national information through the joint development of reliable, practical methods and tools, for example the World Health Survey; the CHOICE initiative (Choosing interventions that are cost-effective); tools for monitoring functions and sub-national health-system performance.

Policy options for health-system financing and human resources

WHO is building a more evidence-based understanding of policy options in health financing and human resources across the Organization.

(ii) Country support and capacity-building

Building national skills in the generation of information

Continuing the strategies mentioned above. There will also be more effort to develop local networks to provide country to country support.

Strengthening national health-information systems

Several countries wish to link baseline assessment with efforts to improve information systems. WHO proposes to take an information needs-oriented view of information system development, and review how to combine strategies such as sentinel surveillance and periodic surveys with routine facility-based reporting systems.

National capacity to use evidence for policy and management

WHO is developing a variety of strategies to build skills in policy analysis and development: national health policy reviews; the Management Effectiveness Programme; rapid health system assessments.

(iii) Expanding WHO inter-country networks on health systems

There is a need for a more systematic approach to facilitating cross-country support in assessment and policy analysis. Existing global and regional networks include the Regional Observatories; the national health accounts partnerships; the Global Alliance for Health Policy and Systems Research; and professional and provider networks. Where appropriate and needed, these will be more systematically strengthened. EHSPi will evolve into a network that brings together those that generate and those that use evidence.

4. SPRG comments and recommendations

SPRG noted the experience of UNICEF which presents a ranking of the performance of individual countries in its publication 'Progress of Nations'. In this publication, UNICEF reviews broad issues affecting child health and welfare but it also includes tables showing specific health achievements of individual countries in relation to their Gross National Product (GNP). From the analysis, UNICEF presents a measure called "National Performance Gap" (NPG) which is derived from the observed health indicator compared with the predicted level on the basis of the nation's GNP. This analysis has been presented for such indicators as the Under-Five Mortality Rate (U5MR), the Maternal Mortality Ratio (MMR), childhood malnutrition, etc. For example, in the case of U5MR, the national performance gap in a particular country is the difference between the actual level of U5MR and the expected level. The expected level of performance for U5MR is derived by fitting a curve to country data represented by points on a graph whose axes are GNP per capita and U5MR. The curve is fitted to match the overall shape of the country data points, using a least-squares regression method. The expected level of performance is the level predicted by the regression line for each level of GNP per capita. The NPG enables each country to assess its performance relative to its level of national income. It draws the attention of a country that is performing worse than predicted according to its GNP per capita. This gap serves to highlight problems that need special attention.

Health authorities find such ranking that is based on clearly defined objective criteria easy to understand and acceptable.

Members of SPRG observed that when moving from diagnosis to policy formulation, policy makers can be faced with an overwhelming amount of information, and ways of showing the potential effects of different policy options would be useful. It was also noted in SPRG that much discussion has focused on WHO reaching top policy makers, but there are also national responsibilities in increasing the use of evidence, by orienting technocrats with managerial responsibilities.

SPRG emphasized that WHO needs to ensure that the public understands the key messages from health systems performance assessment. It will be essential for WHO to think of how to handle public relations at the global as well as national levels for the next Report on HSPA. Access to information on health systems will also be improved through the wider dissemination of country-specific analyses both in electronic and printed form.

SPRG noted that WHO will be unable to meet all demands for direct country support on health systems.

In the Summary Document WHO states that it hopes EHSPi would provide a platform to ensure the policy relevance of HSPA, and to develop national capacities for monitoring and improving performance. The initiative could 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.

SPRG recommends that WHO continue exploring this approach as a vehicle for constructive engagement with countries on health-system performance and ways of improving it. It should also collaborate with countries in the development of practical methods and tools, and provide opportunities to strengthen national capacity in conducting analysis of the performance of national institutions and programmes within the health system. EHSPi will also be of value to other stakeholders in the health field. Whilst the primary focus should be on working with governments, WHO should ensure that other relevant stakeholders are informed and involved. A broad programme of technical assistance based upon the EHSPi experience should be considered.

In terms of increasing policy relevance, it is important that WHO develop indicators of the different health system functions. One can envisage a 'core' group of indicators that could be used in every country, which would facilitate comparisons of performance between health systems. Another more detailed set of function indicators could provide a menu from which Member States can select additional items. SPRG suggests that some basic principles be observed during the development of indicators of health-system functions. The indicators should:

- be policy relevant;
- be easy to use and to understand;
- be sensitive to changes in both directions;
- provide clues about the factors influencing level and change, especially those within the purview of the health system;
- be sustainable, i.e. affordable, reliably collected, and within the capacity of host countries to produce;
- be compatible with local culture and social systems.

5. References

- Almeida, C. M., P. Braveman, M. R. Gold, C. L. Szwarcwald, J. M. Ribeiro, A. Miglionico, J. S. Millar, S. Porto, N. R. Costa, V. O. Rubio, M. Segall, B. Starfield, C. Travassos, A. Ugá, J. Valente, and F. Viacava (2001): Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357(9269): 1692-1697.
- Appleby, J. and A. Street (2001): Health system goals: life, death, and...football. *Journal of Health Services Research and Policy*, 6(4): 220-225.
- Braveman, P., B. Starfield, and H. J. Geiger (2001): World Health Report 2000: How it removes equity from the agenda for public health monitoring and policy. *British Medical Journal*, 323(7314): 678-681.
- Kvale, G. (2000): Inequalities in health. Feedback. *Bulletin of the World Health Organization*, 78(6): 856.
- McKee, M. (2001): Measuring the efficiency of health systems. *British Medical Journal*, 323(7308): 295-296.
- Makinen, M., H. Waters, M. Rauch, N. Almagambetova, R. Bitran, L. Gilson, D. McIntyre, S. Pannarunothai, A. L. Prieto, G. Ubilla, and S. Ram (2000): Inequalities in health care use and expenditures: empirical data from eight developing countries and countries in transition. *Bulletin of the World Health Organization*, 78(1): 55-65.
- Navarro, V. (2001): World Health Report 2000: Response to Murray and Frenk. *Lancet*, 357(9269): 1701-1702.
- Navarro, V. (2002): The World Health Report 2000: Can health care systems be compared using a single measure of performance? *American Journal of Public Health*, 92(1): 31-34.
- Nord, E. (2002): Measures of goal attainment and performance: A brief, critical consumer guide. *Health Policy*, 59(3): 183-191.
- Rosén, M. (2001): Can the WHO Health Report improve the performance of health systems? *Scandinavian Journal of Public Health*, 29(1): 76-80.
- Van der Stuyft, P. and J. P. Unger (2000): Improving the performance of health systems: The World Health Report as go-between for scientific evidence and ideological discourse. *Tropical Medicine and International Health*, 5(10): 675-677.

Walt, G. and A. Mills (2001): World Health Report 2000: Comments. *Lancet*, 357(9269): 1702-1703.

Wibulpolprasert, S. and V. Tangcharoensathien (2001): Health systems performance: What's next?. *Bulletin of the World Health Organization*, 79(6): 489.

WHO Regional Office for the Americas (2001): Regional consultation of the Americas on Health Systems Performance Assessment: Final Report. 8-5-2001, Washington, D.C.

WHO Regional Office for Africa (2001): General Report on the Regional Consultative Meeting on Health Systems Performance Assessment: Final Report. 18-5-2001, Harare, Zimbabwe.

WHO Regional Office for the Eastern Mediterranean (2001): Report on the Regional Consultation on the Conceptual Framework for Health System Performance Assessment. 9-7-2001, Ain Saadeh, Lebanon.

WHO Regional Office for Europe (2001): Report of the Regional Consultation on Health Systems Performance Assessment. 3-9-2001, Copenhagen.

WHO Regional Office for South-East Asia (2001): Report of the regional consultation and technical workshop on health systems performance assessment. 18-6-2001, New Delhi, India.

**MEMBERS OF THE
SCIENTIFIC PEER REVIEW GROUP
WITH INSTITUTIONAL AFFILIATIONS**

Walid Ammar

Director-General
Ministry of Public Health
Lebanon

Sudhir Anand, Chair

Professor of Economics
University of Oxford
United Kingdom

Katarzyna Kissimova-Skarbek

Krakow School of Public Health
Poland

John Eisenberg*

Director
Agency for Healthcare Research and Quality (AHRQ)
United States of America

Timothy Evans

Director, Health Equity
The Rockefeller Foundation
United States of America

Toshihiko Hasegawa

Director, Department of Health Care Policy
National Institute of Health Services Managements
Japan

Ana Langer

Regional Director, Latin America and the Caribbean
Population Council
Mexico

Adetokunbo O. Lucas

Adjunct Professor of International Health
Harvard School of Public Health
United States of America

Lindiwe Makubalo

Chief Director, Epidemiology Research & Evaluation
Department of Health
South Africa

Alireza Marandi

Chairman of the Board of Trustees
Chairman of the Board of Directors
Breastfeeding Promotion Society
Iran

Gregg Meyer**

Director, Center for Quality Measurement and Improvement
Agency for Healthcare Research and Quality (AHRQ)
United States of America

Andrew Podger***

The Secretary
Department of Health and Aged Care
Australia

Peter Smith

Professor of Economics
Centre for Health Economics (CHE)
University of York
United Kingdom

Suwit Wibulpolprasert

Ministry of Public Health
Thailand

- * John Eisenberg regrettably passed away on 10 March 2002
- ** Gregg Meyer joined SPRG in March 2002
- *** Andrew Podger was appointed Public Services Commissioner, Commonwealth of Australia, shortly before the final meeting of SPRG

WHO MEETINGS AND CONSULTATIONS ON HEALTH SYSTEM PERFORMANCE ASSESSMENT

1. Conference on Summary Measures of Population Health, Marrakech, Morocco, 6-9 December 1999
2. 1st Preparatory Working Group Meeting on Measuring Health Status, Geneva, Switzerland, 2-3 August 2000
3. 2nd Preparatory Working Group Meeting on Measuring Health Status, Geneva, Switzerland, 4-5 September 2000
4. Meeting of Committee of Experts on Measurement and Classification for Health, Geneva, Switzerland, 11-12 September 2000
5. Joint ECE/WHO Expert Meeting on Measuring Health Status, Ottawa, Canada, 23-25 October 2000
6. Meeting on Health Systems Performance Measurement, New Orleans, USA, 8 January 2001
7. Regional Consultation of the Americas on Health Systems Performance Assessment, WHO/AMRO, Washington, USA, 8-10 May 2001
8. Regional Consultation and Technical Workshop on Health Systems Performance Assessment, WHO/SEARO, New Delhi, India, 18-21 June 2001
9. Regional Consultation on Health System Performance, WHO/WPRO, Manila, Philippines, 3-5 July 2001
10. Regional Consultation on the Conceptual Framework for Health System Performance Assessment, WHO/EMRO, Ain Saadeh, Lebanon, 9-11 July 2001
11. Regional Consultative Meeting on Health Systems Performance Assessment, WHO/AFRO, Harare, Zimbabwe, 18-20 July 2001
12. Technical Consultation on Effective Coverage in Health Systems, Rio de Janeiro, Brazil, 27-29 August 2001
13. Regional Consultation on Health Systems Performance Assessment, WHO/EURO, Copenhagen, Denmark, 3-4 September 2001
14. Meeting on the Stewardship Function in health systems, Geneva, Switzerland, 10-11 September 2001
15. Meeting on Responsiveness Concepts and Measurement, Geneva, Switzerland, 13-14 September 2001
16. Meeting on statistical methods for enhancing the cross-population comparability of survey results, Cambridge, Massachusetts, USA, 1-2 October 2001
17. Meeting on Fairness of Financial Contribution, Geneva, Switzerland, 4-5 October 2001
18. Technical Consultation on the Measurement of Health Inequalities, Geneva, Switzerland, 7-8 November 2001