Judging Health Systems: Reflections on WHO's Methods

Philip Musgrove (*)11 November 2002

Summary

The "attainment" indicator values in WHO's World Health Report 2000, whatever their conceptual merits, are spurious: only 39 percent are based on actual country-level observations of appropriate variables-only 18.5 percent, for indicators other than disability-adjusted life expectancy. 61 percent are imputations from regressions based on real data for a few countries. Furthermore, the "responsiveness" indicators are not comparable across countries; and in three cases the values obtained from expert informants were discarded in favour of imputed values, without persuasive justification. The indexes of composite attainment and "performance" for ranking countries are also meaningless, because they are based on these imputations. WHO member governments were not informed of the methods and sometimes suffered unjust criticism over the published rankings. If judgements about performance are worth pursuing, they should be based on real data, represent methodological consensus among governments, be built up from assessments at much less aggregated levels, and be useful for public policy.

By way of explanation

Readers of the World Health Organisation's World Health Report 2000 [1] will have noticed that although many numbers are presented in the ten Annex tables, there is scant reference in the text to the indicators used to describe "attainments" or "performance", or their implications for health systems. One critic, remarking on this, says that "Fortunately,...the report appears to make very little connection between the results of the performance analysis and the implications for undertaking [the] functions [of health systems]." [2]. Part of the reason is that the text authors were told essentially nothing about how some of the indicators were estimated until near the end of production of the Report, and the Annex was sent to the printer some time after the text.

Two exceptions to this separation should be noted. First, I participated in many of the discussions concerning the "fair financing" indicator and co-authored two WHO Discussion Papers describing its logic and construction [3-4]. Second, Chapter 2 of the *Report*, which I wrote, offers brief explanations of all the different indicators, with somewhat more detail being provided in the Explanatory Notes to the Annex. Chapter 2 was the last to be written, and reflects what I learned from several of the team working on the Annex, who are listed under "Working Groups" on p. 206 of the *Report*, in the last weeks of text production.

(*) During the period 1 September 1999 to 22 August 2001 the author was seconded from the World Bank to the World Health Organisation, where he served as editor-in-chief of *The World Health Report 2000—Health Systems: Improving Performance.* The interpretations and conclusions expressed in this paper are entirely those of the author. They do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent. Nor do they necessarily represent those of the World Health Organisation, its Executive Directors, or its Member States. I am grateful to two anonymous referees and several other readers for extremely helpful comments.

What follows concerns the numbers in the Annex, and how they were made, not the text. Nothing said here is meant to denigrate the effort mounted by a large team of people to develop the indicators and produce the numbers. Nor do I attribute my views to the co-authors of the text, who were less involved than I with these issues. References to "WHO decision-makers" mean Dr. Christopher Murray and Dr. Julio Frenk, with whom I worked directly, and who may be presumed to have had the backing of the Director General. References to "WHO staff" include a larger number of people who worked on the Report but did not have authority to make decisions about its content. I do not know the views on what is discussed here of several members of the WHO Cabinet or staff in other management positions, and the opinions I have heard from some of them are not documented. The standard caveatthat unless stated otherwise, when opinions are expressed they are those of the author and not necessarily those of anyone else—applies here with even more than the usual force.

The numbers in the Annex

Tables 2, 3 and 4 of the Annex present indicators that WHO has previously published, so they are not discussed in what follows. Annex Table 8 presents estimates of several variables from national health accounts; these also are not considered here. The attainment indicators are presented in Annex Tables 5 (health status and health inequality), 6 (responsiveness) and 7 (fairness in financing). Annex Table 9 presents WHO's overall "attainment" measure, which is a linear combination of the five indicator values; Annex Table 10 converts these into the WHO "performance" indexes (one for health status alone and one for the composite attainment); and Annex Table 1 presents the rankings of countries that emerge from these exercises.

The principal feature of the attainment numbers is that most of the values were not derived from any detailed national-level information. This was not for lack of trying: WHO staff mounted an enormous effort to obtain such information. But data corresponding to the chosen indicators were simply not available for many countries, or not in time to include. WHO decision-makers therefore chose to run linear regressions on the estimates based on real data, and use the results to impute values for countries for which there were no such data. The imputed values are indicated by italicised numbers in the Annex tables, but the italic typeface is so close to vertically aligned that several people to whom I have shown the tables and asked for reactions, could not see the difference between the real and the imputed values until it was pointed out, or their attention was drawn to the footnote which says only that "Figures in italics are based on estimates". This does not explain anything about the estimates or distinguish them from the numbers based on disaggregated country-level information, which also always involved estimation. In the Explanatory Notes to the Annex, the footnote is amplified to say that the indicator in question "has been estimated using indirect techniques and information on important covariates", but the techniques are not explained and only some of the covariates are named. In any case, the distinction between measured and imputed values

disappears in Annex Tables 1, 9 and 10, where the numbers are presented as if they were all on the same footing. Even for the five countries (Bangladesh, Brazil, Nepal, Peru and Thailand) for which all the five indicators were measured directly, the performance index—which depends on estimating a "frontier" based on the values for all the other countries—is therefore the result of imputations.

Because it was evident by the time Chapter 2 of the Report was written that a large share of the "data" were going to be imputed, I decided that most of the material presented in that chapter to explain the indicators would draw only on countries for which such detailed information had actually been analysed. This is the case for the illustration of the distribution of life expectancy at birth (Figure 2.3) and of household health financing contributions (Figure 2.5). In the case of the "responsiveness" indicators (Figure 2.4), comparisons among countries were avoided altogether, and the indicators for the seven different components were instead compared within each of 13 countries. The only instances of comparisons across countries that include the imputed values are Figures 2.6 and 2.7 (based on Annex Tables 9 and 10), and the accompanying discussion. I now believe it was a mistake to have put them in the text, which otherwise had scrupulously omitted such presentations.

The result of this way of making the numbers was that the Ministers of Health of the world may well have felt, on the day the Report was published, that they were in the position of parents whose children had been given grades in courses in which they did not know they were enrolled. Several WHO Representatives and Liaison Officers were also taken by surprise. As some of them explained to me, the advance copies of the Report and press materials they were given immediately before publication could not help them explain to outraged or baffled government officials where the numbers came from. Only 39 percent of the indicator values represent real data, and that share falls to 18.5 percent if one sets aside the estimates of disabilityadjusted life expectancy. This was the only indicator not imputed by simple regression-although much imputation went into the estimate-for 118 of the 191 Member States of WHO. Among those 118 countries are 25 that appear in the top 30 positions of the ranking by attainment and 23 that appear in the top 30 positions ranked by overall performance. Note 1, below, summarises the amount of imputation by country and by indicator.

The particular case of "Responsiveness"

This indicator differs from the others in being estimated for the Report 2000 by an entirely different method than WHO decision-makers intended to use in future exercises. For the first attempt, groups of key informants were recruited in each of 35 countries and answered a questionnaire about their own country's health system. The heads of the key informant groups were agreed, however, that the results cannot be used to compare one country with another, because no informant actually looked at, or was likely to know much about, any country but his or her own. When these heads of groups met in Geneva in December 1999, they expressly urged WHO not to use the numbers for any international comparisons, because, as one of at least four written protests put it, "There was a unanimous agreement that the instrument was unsuitable in capturing information universally on the domains that were decided by the WHO. We were made to understand that this was a pilot study and findings of this attempt would enable identified issues to be included in some final survey with a representative sample of adequate

number. But ranking countries based on this pilot study has been inappropriate and embarrassing." [5]. WHO staff did not comply with that request.

Even so, the numbers were not always respected. In five instances, imputed values were published for the responsiveness indicators although actual responses were obtained from key informants. In two cases the key informants gave their opinions on only one province (Shandong in China) or state (Andra Pradesh in India) rather than the entire country. It is certainly arguable that one part of such a vast and varied country may not be representative of the whole. Nonetheless, it surely represents the country better than an estimate based on opinions about 30 other countries, by local informants, with the province or state treated as if it were just one more country for purposes of estimation.

In three other cases, the key informants' opinions were disregarded, and the imputed values used instead, without the excuse of incomplete geographic information. The three were Chile (rating improved), Mexico (rating lowered) and Sri Lanka (rating improved). WHO Discussion Papers 21 [6, pp. 22, 23 and 25] and 22 [7, p. 9] give what are supposed to be reasons for replacing the key informants' evaluations. In two cases the justification given is the existence of a health reform, and in the other case it is a civil war. Neither of these phenomena is limited to the countries named, nor is there any explanation of why war or reform would make an imputed value more accurate than the opinion of well-informed observers from the country. Still less is it clear why a health reform in progress would make such observers err in one direction in Chile but in the opposite direction in Mexico.

When I discovered this substitution, I wrote to Dr. Murray, who had authorised the changes, that "if that doesn't qualify as manipulating the data, I don't know what does...At the very least, it gravely undermines the claim to be honest with the data and to report what we actually find." [8]. In reply, Dr. Murray said that "if results from any survey lack face validity it would be rather counterproductive to simply go with them. It is the careful interplay of informed assessment of the quality of the results and empirical findings that is the hallmark of the development of good data systems" [9]. I leave the reader to judge the "face validity" of that justification and of the explanations for each country.

I regard these as not simply statistical or even political issues, but ethical ones. WHO insists, and rightly, that when member governments send it data that the Organisation needs for its work, they should not misrepresent the status, nature or meaning of that information. It is equally important that WHO publications meet the same standards. If its valuable imprimatur is to be respected WHO must, like Caesar's wife, be above suspicion. My efforts to persuade Drs. Frenk and Murray that the publication of these numbers was unethical were, I am sorry to say, in vain.

The question of whether the numbers are honest has an occasional unintended comic aspect. The Russian Minister of Health, without knowing how the numbers were arrived at to rank Russia 130th among world health systems, declared himself unperturbed by WHO's judgement, believing it to be based on correct numbers, unlike the practice in his country until recently, and to be an honest reflection of the situation ("'nullement vex [sic]' par l'appréciation de l'OMS sur le système de santé de son pays, estimant qu'elle était basée sur des chiffres exacts, 'contrairement a la pratique dans notre pays il n'y a pas si longtemps', et qu'elle était 'le reflet honnête' de la situation") [10].

The use and interpretation of imputation techniques

The presentation of the results included estimates of "uncertainty intervals" around the numbers, which were carried over into the attainment and performance indexes. This appears to add a welcome element of modesty to the estimates. When the estimates come from regressions, however, the standard errors of the coefficients and the resulting standard error of estimate only show the likelihood that the true value falls into the interval indicated, with the probability chosen, provided the regression model is correct. Whatever is left out of the model must be random, normally distributed noise, and not one or more variables that would give entirely different imputed values.

Imputation can lead to utterly misleading results: a clear example occurs with the equation for the fair financing indicator [11]. The Gini coefficient on income is among the explanatory variables, on the reasonable supposition that the more unequal income is in a country, the more likely it is that households will have very unequal health spending, particularly out-of-pocket spending, relative to their capacity to pay. The estimation did not show the Gini coefficient to be statistically significant, but it would not matter if the error were smaller and the effect more significant. The regression suggests that high income inequality makes it hard for a country to achieve fairness (on the WHO definition) in paying for health care, which is not at all surprising. But to judge how well a health system performs, one does not want to know that-one wants to know how well the system offsets or compensates for that handicap, by the way it actually finances health. A country with a high Gini coefficient that nonetheless achieves a relatively good distribution of the finance burden among households deserves praise for overcoming the income inequality handicap, rather than being penalised as if the health system had simply mirrored other forms of inequality in society. Imputed values can of course err in the opposite direction, and by giving a false sensation of excellence, undermine the necessary and urgent efforts to solve the crisis in a health system ("causan perjuicio al dar una falsa sensación de excelencia y, por lo tanto, tienden a entorpecer los necesarios y urgentes esfuerzos para solucionar la crisis") [12].

Finally, when a number of imputations are combined into some overall index, they must be clearly interpretable. The child survival inequality and financial fairness indicators used single estimating equations with only three independent variables each, which are easy to interpret. There are however seven components for the level of responsiveness, with an equation for each one, and an eighth equation for the distribution. When the equations used to impute these values are added up, the result is not interpretable. The same variable sometimes pulls in different directions and enters in different forms (linear, logarithmic, exponential, and power, sometimes alone and sometimes interacted with one or more other variables) in the several equations, and it is impossible to make sense of any overall effect.

When a considerable number of the governments composing WHO's Executive Board asked for more complete technical explanations in time for their meeting in January 2001, they were given a collection of edited versions of existing Discussion Papers, which explains much of the methods and the research behind them, plus some new material [13]. However, nothing was included to explain the imputations.

Usefulness for health policy

Because of their deficiencies, the Report's overall attainment and performance estimates are of no use for judging how well a health system performs-either absolutely or compared to other countries. In fact, they illustrate the mathematical truth that the difference between two complex numbers may be entirely imaginary. The problems of data and imputation invalidate even comparisons on a single indicator, unless they are limited to the countries for which there were real data-and not even then, for the responsiveness indicators. The composite attainment measure is therefore spurious, as is the performance or efficiency index, quite apart from the substantial further assumptions and estimations it involves. This has not stopped WHO staff from attempting to "explain" differences in countries' performance, as though the performance itself were real and accurately estimated [16]. The explanatory or determinant variables proposed are plausible, and may help understand why some countries have better health than others. But to undertake such analysis prematurely violates a basic rule of scientific inquiry, which is not to explain a phenomenon until it is confirmed to exist. To do otherwise amounts to guessing who built the canals on Mars.

What the WHO exercise has achieved is to provide some information—much less than has been claimed, because of all the data shortcomings, but still something that may be of interest or use. What it has yet to provide is a convincing demonstration that the information constitutes evidence for assessing or guiding health policy. This is partly because of the conceptual flaws in the exercise, but also because even a good measure of relative performance would be of no use to any government unless it suggested how to improve that performance.

Showing governments (and other interested parties) how to improve performance is not the same thing as "explaining" performance by further statistical research. The regression equations used to impute values of the indicators are a clear illustration: to discover that richer countries have more responsive systems is of little help to a poor country, nor is much gained from knowing that inequality in child mortality is lower-as it virtually has to be-when total child deaths are fewer. What a concerned government wants to know is what it can do about systemic failings. The effort to rank countries rather than to devote the same resources to helping them do better, looks like a distraction rather than a contribution. What WHO has accomplished to date in helping its member governments-eradicating smallpox, nearly eliminating polio, controlling onchocerciasis, combating tuberculosis, AIDS and smoking-related disease, and providing scientific and practical help to development banks and donor agencies as well as governments-has always depended on real and specific knowledge, not on highly aggregated and questionable assessments.

The extreme level of aggregation is one reason the results are of no use for policy. The fact that so many of the component numbers were imputed, and sometimes manipulated rather than measured, is another. A third reason, no less important, is that WHO decision-makers avoided any real consultation or participation by governments until forced to admit them. There were two consultative meetings, in December 1999 and January 2000, attended by academic experts and by some staff from WHO headquarters and the Regional Offices. I attended both of them, and can vouch that while the WHO framework was well explained and the indicators were described in general terms, participants were not told exactly how any of the numbers were being calculated, nor

about the intention to publish imputed values when there were no data. So far as I can determine, nothing said by participants at either meeting changed anything about the methods or the numbers. Subsequent requests for more information, even from Ministers of Health, were not fully and promptly answered [17]. WHO staff have insisted that "We should not underestimate the intelligence of policy makers" who can interpret the published measures for themselves [18], but in my view, the Organisation did not respect the intelligence of those same policy makers when defining and calculating the indicators and then failing to explain them adequately.

Dr. Murray also exaggerated the novelty of ideas and the indicators in the Report. This went so far that he accused Professor Hsiao of Harvard University of plagiarising the main ideas of the WHO framework in a paper that also discussed health system goals [19], necessitating a refutation by Hsiao with reference to his publications predating the Report and differing with it on substantial points [20].

Finally, the usefulness of whatever WHO does depends not only on the content, but on relations with member governments. Sometimes the Organisation must pick quarrels with governments and criticise their policies or their efforts: "good" relations are not always the most productive. On those occasions, criticism must be based on solid scientific knowledge. It is an entirely different matter to publish incomplete or spurious information which can facilitate unjustified attacks by a government's domestic critics.

Why did WHO decision-makers proceed as they did ?

Given the scientific and ethical objections, to say nothing of the political risks involved, the question naturally arises why WHO took the course it did. Fortunately it is not necessary to speculate. I was told in several conversations with Dr. Frenk or Dr. Murray or both, sometimes in the presence of others, why they thought WHO had to publish an index of health sector performance, even with no consultation with the governments and hardly any with other WHO staff, and with most of the numbers imputed rather than actually measured. These reasons included—

The favourable opinion of Amartya Sen, who led the creation of the Human Development Index (HDI). With all due respect for Professor Sen, that Index is simply a combination (with some transformations) of existing indicators received from governments. That distinguishes it from the combined WHO attainment measure that includes newly-designed indicators not provided by or even discussed with governments. The HDI also does not pretend to establish a frontier of what countries ought to be able to achieve, so while it suffers from some of the same deficiencies as the WHO index, in that respect it is quite different.

The supposition that no one was adequately concerned with health system problems, so an index, including a ranking, was needed to call attention to them. This argument gives no credit to all those, in or out of government or other international organisations, who have been working on health system reform and improvement in many countries and for many years.

The assertion that nobody would have paid attention to a partial analysis limited to real data, no matter how interesting the relations. It may be true that a more limited and evidence-based analysis would have attracted less notice, because policy-makers pay too much attention to scorecards and too little to what would really help them do their job. However, that matters only if the objective was to gain attention. Besides, if Ministers of Health and others do care excessively about rankings, that is partly because WHO (and other international organisations) push them to do so.

The claim that the estimates, imputations, shortcuts, etc. used to fill in the tables were better than any previous ones, such as those used in the World Bank's World Development Report 1993 [14] or WHO's own estimates of the Global Burden of Disease [15]. The argument was also that it makes no difference what use the results are put to—ranking or rating is no different from just estimating.

Finally, the claim that WHO had to produce rankings in time for the 2000 Report and could not wait for better data or a more consultative process of collection and interpretation. Any such urgency was entirely selfimposed; it did not arise from the nature of the exercise or from the needs or wishes of the Member States.

Rethinking the "frontier of the possible"

Clearly, there is a need for a sharp line between actually measured indicators and imputations, with none of the latter used as if they were real measurements. This is the only scientific way to proceed, and is moreover consistent with WHO's stated aim of eventually quantifying, with real disaggregated country-level data, any attainment indicators it uses. At least some of the indicators themselves also need to be re-thought. Just as clearly, a political and attitudinal change is needed to restore WHO's credibility and repair relations with member governments. Supposing the indicators to be conceptually satisfactory, adequately discussed beforehand with governments and academic experts, and properly measured from the appropriate country-level data, the question remains whether they can yield meaningful judgements or rankings of countries' performance.

The attempt to measure performance in the Report depends on assuming a production function for health (or other attainments) which at any moment is essentially uniform and available to all health systems. Compared to making steel or growing rice, "producing health" is so much more complicated as to call into question the idea that there is a common underlying function at all. In the published estimate derived from such a function, the most striking result is that the performance scores are correlated with health expenditure per capita [1, Figure 2.6]. At expenditures below US\$ 100, about half the countries receive scores of 50 % or less, whereas such poor performance is rare at higher spending levels. Since the numbers are spurious, this does not imply anything about a minimum or threshold level of spending necessary for a well-functioning health system, despite efforts by WHO staff to interpret it that way [16, pp. 4-5]. But when a supposed frontier of what is attainable passes through a relatively large area where no observations fall close to it, something is wrong. There is no assurance that the countries in that area actually could get closer to the frontier; the prima facie evidence is that the frontier has been drawn too far away from them.

Most of the countries that seem to perform poorly when the frontier is drawn high overhead—all of the lowestranked 18, and 33 out of the lowest-ranking 36 countries are in sub-Saharan Africa. The Report says that "A large part of the explanation is the HIV/AIDS epidemic" [1, p. 43]. That AIDS is responsible for a large share of deaths and of the total burden of disease in those countries is not in doubt [1, Annex Tables 3 and 4]. There is no way to distinguish from the data, however, between two very different interpretations: that AIDS is making it hard to reach the (fixed) frontier, or that it has moved the frontier itself downward, reducing the best attainable health status. For political and policy purposes there is an enormous difference between these views: have the African countries fallen toward the floor, or has the roof collapsed on them ?

Choosing the second interpretation means recognising that deaths and social disruption from AIDS have impaired African societies' capacities even to deal with the burden from other, pre-existing diseases. Under the impact of the epidemic, money and education—the two variables defining the frontier—do not mean what they used to mean; they have become less effective in improving health. In contrast, the first interpretation claims that African governments could and should control the health damage from AIDS, with the existing levels of expenditure and schooling, and should be held responsible for not doing so.

What is missing from this thinking is any sense of the feasibility of controlling AIDS. There is now considerable evidence that some preventive interventions, in some countries, have contributed to slowing or even reversing the trend toward higher AIDS incidence [21]; but because the response depends so much on individual attitudes and behaviour, which vary from one culture to another, there is still nothing like a guarantee that control is feasible everywhere with known interventions. Beyond the issue of what interventions would be required for control, is the question of how much they would cost. Recent estimates suggest that for sub-Saharan Africa alone, prevention, care and antiretroviral therapy for AIDS would require incremental spending, respectively, of at least 1.17, 1.05 and 0.72 billion dollars annually by the year 2007, and two, four and eight times those amounts by 2015 [22, Tables A.9-A.11], just to keep prevalence from increasing in low-prevalence countries, and perhaps to start lowering it in high-prevalence countries. That conclusion is equivalent to saying that the frontier has moved down because of AIDS.

What money can accomplish depends on knowledge, and that can change suddenly when the causes of ill health change. What one can do with money today also depends on what was done with the money of yesterday and the day before-money that at best is crystallised in people, buildings, etc. which can produce health today and tomorrow. (At worst, past money is irretrievably lost through bad investments, or the real resources it purchased have been fully consumed or depreciated.) If one country or health system always spent resources wisely in the past and another did not, equality of their current expenditure does not imply equality of potential attainment in the short run. More generally, two countries with the same values of schooling, health expenditure and other variables, but very different recent histories, will not have identical possibilities for future progress. In particular, a country that attained better health in the recent past should find it easier to regain that level-barring some disastrous change-than it would be for another country to reach that level for the first time. Basing estimates of the frontier on information from several recent years rather than just one, as the Report does, only partly corrects for this historical path-dependence.

Extra money (or any other incremental resource such as knowledge) takes time to become effective. The time required differs among diseases, because scaling up control efforts requires more in the way of investment in infrastructure, training or other inputs for some than for others [22, p. 86]. In consequence, a frontier of the possible cannot be defined independently of the composition of its disease burden and the interval in which a country is supposed to improve its health system. The notion that there is a health system which actually controls the use of resources and could apply them very differently also becomes questionable when most spending is out of pocket and there are few central decisions about anything.

Where to go from here ?

The foregoing discussion suggests that performance measurement relative to a frontier of what health systems should be able to achieve is a chimera, at least in the highly aggregated, top-down fashion that WHO decision-makers have pursued. It is still valuable to measure attainments of various kinds, and to look for ways that health systems can achieve more and use resources better. Breaking down "performance" along one or more dimensions seems the right direction to go.

WHO staff recognise the limitations of treating an entire large country as a homogeneous unit, and therefore propose to carry performance assessment down to subnational levels where data permit. If this is done the problem of attribution—of what to hold the health system responsible for—will get worse, because results in one state, province or area will depend on factors operating at the national level and possibly also in other sub-national areas. Tracing out these connections presents a major challenge. The problem will also get better, to the extent that different geographic areas really have distinct health systems whose differences need to be taken into account.

Spatial disaggregation is welcome, but functional separation is surely more important for policy. That means, for example, evaluating how the hospital subsystem functions and how it contributes to health system performance as a whole. Determining which outcomes to hold primary care, or hospitals, or other sub-systems responsible for is very complex: hospital outcomes may look bad because the primary care sub-system does not do its job properly. Getting those assessments right is of much more value than overall judgements, because it directs attention to specific areas, organisations and policies. An ideal method of performance assessment would then build up from inputs and outcomes at individual facilities or programs to more aggregated analyses. If a proper way to do that can be determined, outcomes at one level could be treated as intake at a higher level, and achievements could be traced from the bottom up, whether or not they ended with any measure of overall system performance. If that cannot be done in a reasonable time and at reasonable cost, then it is better to abandon the performance measurement exercise and devote the resources to uses of more immediate value to the people-patients, taxpayers and consumers-for whose benefit health systems exist.

Conflict of interest statement

As stated at the outset, this article derives from my experience as editor-inchief of the World Health Report 2000, while employed at WHO on loan from the World Bank. The paper was written after leaving WHO to return to the Bank. There is no other possible conflict of interest.

Funding source

No funding for writing this article was provided by any source.

References

- 1 World Health Organisation, World Health Report 2000 Health Systems: Improving Performance. Geneva: WHO, 2000.
- 2 Institute for Health Sector Development (IHSD), "Improving health systems by measuring health status: is WHO serious ?" London: IHSD, September 2000.
- 3 C. J. L. Murray, F. Knaul, P. Musgrove, K. Xu and K. Kawabata, "Fairness in financial contribution: conceptual framework". Originally WHO/GPE Discussion Paper No. 24, published as Chapter 16 of [13].
- 4 C. J. L. Murray, F. Knaul, P. Musgrove, K. Xu and K. Kawabata,

"Estimating health financing contribution from survey data". This is an excerpt from what was originally WHO/GPE Discussion Paper 25, published as Chapter 17 of [13].

- 5 Comments by key informants from Indonesia, India, Poland and Brazil, Appendices 2.1-2.4 to Ministry of Health Policy Unit, Vietnam, "Comments and suggestions of Vietnam Ministry of Health/health policy unit as regards the World Health Report 2000". Hanoi: 2000. The quotation is from Appendix 2.2, "Protest from India to use responsiveness data".
- 6 N. P. Valentine, A. de Silva and C. J. L. Murray, "Estimating responsiveness level and distribution for 191 countries: methods and results". Geneva: WHO/GPE Discussion Paper No. 22, 2000.
- 7 A. de Silva and N. P. Valentine, "Measuring responsiveness: results of a key informants survey in 35 countries". Geneva: WHO/GPE Discussion Paper No. 21, June 2000.
- 8 P. Musgrove to C. J. L. Murray *et al.*, electronic mail, "Censoring the responsiveness results", 30 August 2000.
- 9 C. J. L. Murray to P. Musgrove *et al.*, electronic mail, reply to [8], 30 August 2000.
- 10 From the daily report of "WHO in the news", a collection of press notices about the Organisation that is circulated to all WHO staff, concerning a visit of the Director-General to the Russian Federation.
- K. Xu *et al.*, [Title not determined]. Geneva: WHO/GPE Discussion Paper No. 26, Table 4—draft, not published.
 Academía Nacional de Medicina (Colombia), "Carta a la Ministra de
- Salud". Bogotá: 11 September 2000.
- 13 C. J. L. Murray and J. Frenk, eds., Health System Performance: Concepts, Measurement and Determinants. Geneva: WHO, January 2001.

- 14 World Bank, World Development Report 1993: Investing in Health. Washington, DC: World Bank, 1993.
- 15 C. J. L. Murray and A. L. Lopez, eds., The Global Burden of Disease: a Comprehensive Assessment of Mortality and Disability from Diseases, Injuries and Risk Factors in 1990 and Projected to 2020. Cambridge, MA: Harvard University Press for WHO and the World Bank, 1996.
- 16 J. A. Lauer, A. Tandon, D. B. Evans and C. J. L. Murray, "Determinants of performance", chapter 23 of [13].
- 17 Letter from Dr. José Serra, Minister of Health of Brazil, to Dr. Brundtland, Director-General of WHO, 5 July 2000.
- 18 C.J.L. Murray, J. Frenk, D. Evans, K. Kawabata, A. Lopez and O. Adams, "Science or marketing at WHO ? A response to Williams", *Health Economics* 10(4), 2001.
- 19 W. C. Hsiao, "A Strategic Framework in Mobilizing Domestic Resources for Health", Commission on Macroeconomics and Health, Working Group 3 paper. Cambridge, MA: draft of 25 October 2000.
- 20 W. C. Hsiao, "Clarification of My Writing and the WHO World Health Report, 2000", unpublished note attached to e-mail letter to WHO, 27 November 2000.
- 21 P. Jha et al., "The evidence base for interventions to prevent HIV infection in low and middle income countries", Commission on Macroeconomics and Health, Working Group 5, Paper 2.
- 22 Commission on Macroeconomics and Health, *Improving Health Outcomes of the Poor: Report of Working Group 5.* Geneva and London: WHO and the London School of Hygiene and Tropical Medicine, 2001.
- 23 E Gakidon, G King. An individual-level approach to health inequality child service in 50 countries. WHO/GPE discussion paper no 18. Geneva: World Health Organization/GPE, 2000.

Note 1: countries for which detailed national data were used to calculate components of the indexes of attainment and of performance, as indicated in the WHR Annex Tables 5, 6 and 7. (The estimates of disability-adjusted life expectancy, DALE, are all based on some national—or at least regional—level data at a minimum, for mortality—rather than projections from regressions.)

Health Inequality (Child Survival)				Responsiveness (Level and Inequality)		Fair Financing (Inequality)	Fair Financing (Inequality)	
56 Countries			_	30 Countries		21 Countries		
Bangladesh	3	Morocco	1	Bangladesh	3	Bangladesh	3	
Benin	1	Mozambique	1	Bolivia	2	Brazil	3	
Bolivia	2	Namibia	1	Botswana	1	Bulgaria	2	
Botswana	1	Nepal	3	Brazil	3	Colombia	2	
Brazil	3	Nicaragua	2	Bulgaria	2	Guyana	1	
Burkina Faso	2	Niger	1	Burkina Faso	2	India	2	
Burundi	1	Nigeria	1	Cyprus	1	Jamaica	1	
Cameroon	1	Norway	1	Ecuador	2	Kyrgyzstan	1	
Central	1	Pakistan	2	Egypt	2	Mexico	2	
African		Paraguay	2	Georgia	1	Nepal	3	
Republic		Peru	3	Ghana	2	Nicaragua	2	
Chile	1	Philippines	2	Guatemala	2	Pakistan	2	
Colombia	2	Poland	2	Hungary	1	Panama	1	
Comoros Is.	1	Rwanda	1	Indonesia	2	Paraguay	2	
Cote d'Ivoire	1	Senegal	2	Korea (Republic)	1	Peru	3	
Dominican	1	Somalia	0	Malaysia	1	Romania	1	
Republic		Sudan	1	Mongolia	1	Russian	1	
Ecuador	2	Tanzania	2	Nepal	3	Federation		
Egypt	2	Thailand	3	Peru	3	Tanzania	2	
Ghana	2	Togo	1	Philippines	2	Thailand	3	
Guatemala	2	Trinidad and	2	Poland	2	Viet Nam	2	
Haiti	1	Tobago		Senegal	2	Zambia	2	
India	2	Tunisia	1	Slovak Republic	1			
Indonesia	2	Uganda	2	South Africa	1			
Japan	1	United	1	Thailand	3			
Kazakhstan	1	Kingdom		Trinidad and	2			
Kenva	1	United States	1	Tobago				
Liberia	1	of America		Uganda	2			
Madagascar	1	Uzbekistan	1	United Arab	1			
Malawi	1	Yemen	1	Emirates	-			
Mali	1	Zambia	2	Viet Nam	2			
Mexico	2	Zimbabwe	2	Zimbabwe	2			
For child survival inequality, Botswana			For responsiveness,		Data on financial			

and Somalia are listed in normal type in Annex Table 5 of the Report, indicating that the index is based on detailed data, but these are typographical errors. Neither country is included in Table 1 of the Discussion Paper, which indicates the sources of the estimates for 56 countries (50 Demographic and Health Surveys, and 8 small-area studies with both sources available for Brazil and for Mexico). For responsiveness, one state in India and one province in China were analysed, but the national estimates were imputed (the state and province, estimates were used in the imputation). The use of imputations for Chile, Mexico and Sri Lanka, despite having actual country estimates, is related in Data on financial contribution were also collected for Germany, but were not used because of the difficulty of estimating expenditure net of insurance reimbursements when the outlay and the repayment occurred in different periods.

The number to the right of each country's name indicates whether detailed data were used for one, two or three of the indicators. (Somalia is given a zero, and Botswana a value of 1, because their estimates of child survival were not calculated from detailed national data). Health inequality (child survival), responsiveness (level and inequality together) and fair financing each account for 25 percent of the composite index of attainment. Therefore the share of that index for a particular country that is based on detailed national-level information is given by 0.25[N + 1], where N is the number above, and 1 is added to account for the share of attainment due to health level measured by disability-adjusted life expectancy (DALE).

the text.

The number of countries for each value of N, and the corresponding share of the information derived from detailed national data, are as follows:

	N=0	1	2	3	Total	
Detailed data share (%) Number of countries	25 118	50 45	75 23	100 5	39 191	

For the 191 Member States of WHO, the mean value of N is $0.555=[(118\times0) + (45\times1) + (23\times2) + (3\times5)]/191$. The corresponding mean share of the composite index derived from detailed national data is 0.389. That is, 61 percent of the numbers which go into the index are derived from imputations and only 39 percent are based on detailed analyses without any projection across countries. There are 118 countries for which N=0, because the only detailed national data refer to health level—that is 62 percent of all the countries, which is only coincidentally almost equal to the 61 percent data share. The mean value of N for indicators other than health level is only 0.234, less than half the overall mean of N. This corresponds to a share of real numbers for those indicators (health inequality, responsiveness and financial fairness) of only 18.5 percent, equivalent to having detailed and complete information for 35 countries.