WORLD HEALTH ORGANIZATION
REGIONAL OFFICE FOR EUROPE

WELTGESUNDHEITSORGANISATION
REGIONALBÜRO FÜR EUROPA

ORGANISATION MONDIALE DE LA SANTÉ
BUREAU RÉGIONAL DE L'EUROPE

???? ????? ??????????? ???????????????
???? ??? ???? ????????????? ?? ??

5031604 2000/6

European Regional Consultation on the World Health Report 2000

Copenhagen, Denmark, 3 - 4 September 2001

The World Health Report 2000: advancing the debate[1]

Martin McKee,
European centre on Health of Societies in Transition
London School of Hygiene and Tropical Medicine
European Observatory on Health Care Systems

Background

In response to concerns voiced after the publication of the 2000 World Health Report (WHR 2000)[i] on assessing the performance of health systems, the Director-General of the World Health Organization (WHO) has decided to establish a process of technical consultation, bringing together personnel and perspectives from different WHO regions, to explore how data on health systems performance might be used in future reports. So far, five of the six regional offices have held consultations. This paper has been prepared to facilitate discussion during the sixth such consultation, to be held by the European Regional Office, in September 2001. The WHO Executive Board has requested that this process should draw on scientific peer review. One element of peer review is the response to published work that appears in the scientific literature.

It is important to stress, at the outset that, given the intense level of controversy that the WHR 2000 has generated and the stage at which the consultation process has reached, ***it is inappropriate for this paper to offer a definitive judgement on the various issues that have arisen***. Instead, it seeks to summarise some of the concerns that have been raised about the report as well as some of the responses to them. It then raises a series of questions that may facilitate discussions during the consultation. This paper should be read in association with the Background Paper for this consultation, prepared by WHO HQ.

While this paper focuses on the methods used in the WHR 2000 to assess performance of health systems, it does not describe them in detail: this has been done elsewhere and the main elements are covered in the accompanying background paper and in the Report itself. In view of the constraints on time for the consultation, it seeks to address some of the general issues that have been raised, avoiding some of the detailed and highly technical methodological issues that have been addressed elsewhere.[ii, iii]

[1] This paper has been prepared for the European Regional Consultation on Health Systems Performance Assessment. WHO Regional Office for Europe Copenhagen 3- 4 September 2001.

The issues

The concerns expressed about WHR 2000 can be considered under the following headings: the underlying philosophy, whether explicitly set out in the text or implied by the selection of indicators; the face validity of the rankings, or whether the relative positions of certain countries are intuitively plausible; the coherence of the performance measures used with the definition of the health system; the approaches used in situations where necessary data were unavailable; technical issues concerning individual measures of performance; issues relating to the overall measurement of performance; and concerns about the breadth of evidence used and the cost of undertaking the exercise in the light of competing priorities.

Underlying political philosophy

Navarro has criticised the philosophy underlying the WHR 2000 on two counts.[iv,v] The first is that the challenges facing health systems can be resolved by "technological-scientific medical bullets … without reference to changes in the social, political, and economic environments in which these problems are produced". This criticism overlaps with issues of attribution of measures of performance to characteristics of health systems and will be discussed later. His second criticism is that the WHR 2000 uncritically adopts a conservative and neo-liberal agenda on the organisation of health systems. Thus, he notes, the report's bibliography draws extensively on views of writers who are active proponents of market-led reforms, while disregarding contrary views. He cites as examples a statement suggesting scope to extend to other countries the reforms of the British National Health Service initiated by Margaret Thatcher, the advocacy of systems that permit money to follow patient flows rather than being determined by prospective budgets, and the expansion of private delivery of health services as a means of enhancing responsiveness. To support this argument he draws attention to the language used in the report, such as "heavy handed state intervention … discredited everywhere", "highly impersonal and inhuman", and "monolithic", which, he contends, are used to draw an analogy between national health services generically (defined as combining public funding and provision) and the system share the same characteristics as the system in the Soviet Union. He then argues that the evidence to support these views is extremely limited and thus the WHR 2000 reflects ideology rather than evidence of effectiveness.

Navarro contends that these views are the latest manifestation of a process that the Pan-American Health Organisation (PAHO) has been engaged in for many years, acting as a means of transmission of a particular American perspective on the role of the state in health care throughout Latin America. In his support he cites praise by PAHO for the privately managed health insurance funds in Chile, which were established under Pinochet and which have been criticised extensively for reducing access to health care among the poor.

Both Navarro and Almeida et al.[vi] criticise the analysis of the failure of the Alma-Ata model of primary care, which the WHR 2000 relates to its internal policy weaknesses and failure to address demand. They contend that external factors have been equally to blame, including economic recession in less developed countries and imposed privatisation policies. They argue that the rejection of this model of primary care reflects a flawed historical analysis and the author's ideological opposition towards what they view as too active a role for government in protecting the vulnerable.

A contrary view that has been cited as showing that the WHR 2000 was actually more balanced was expressed by Helms, writing in the Wall Street Journal.[vii] He argues that, by advocating that the state should have anything more than a token role in the provision of health care, the World Health Organization has adopted a Marxist stance. Instead, he argues, health care should generally be left to the market, with the minimum of state interference.

Murray and Frenk,[viii] in their response to Navarro, have used these polarised views to argue how discussions of health policy become ideological debates, implicitly criticising Navarro's arguments. Apart from this, however, their response to Navarro's criticisms focuses mainly on other issues in his paper, in particular the attribution of levels of health to health systems. In a subsequent response, Navarro argues that his specific criticisms relating to political ideologies have been largely ignored.[ix]

Question:

To what extent are the arguments developed in the text of WHR 2000 based on rigorous empirical evidence and to what extent do they reflect a particular political ideology?

Williams draws attention to, but does not explicitly answer, the question of whether, given that every health care system will reflect its own particular history, dominant ideology, and specific economic and socio-demographic features, it is possible to derive some universal performance criteria.[x] For example, values underlying health care in Europe differ markedly from those in the United States, in terms of issues such as equity, access, and whether health care is a basic human right. Williams does, however, suggest that some half-way position may be appropriate as a negative answer would imply that any comparison would be "at best hard to interpret and at worst totally irrelevant". Leading on from this argument, Williams contends that the focus of the report on overall values is irrelevant to a world in which policy-makers must make trade-offs, and thus choices at the margin.

Question:

To what extent do universal values exist? What are the implications for comparisons of the performance of health systems? For example, the USA is the only industrialised country that fails to ensure virtual universal health coverage for its citizens, and even it has accepted that the state has an important role to play in the funding and provision of health care. Can the views expressed by Helms be rejected as irrelevant in the European context?

Face validity

Several commentators have noted that some of the rankings generated by the process employed in the WHR 2000 are intuitively implausible. Thus, Borowitz (personal communication) has noted that Kazakhstan achieves a higher overall rank than Kyrgyztan even though the reform process in Kazakhstan has collapsed amidst major problems for health care delivery while that in Kyrgyztan is considered by many to be a relative success. Navarro draws attention to the contradiction between the high rankings achieved by Spain and Italy and the poor assessments of their health care systems by their citizens, as judged by international surveys of public opinion.[iv] Williams notes with surprise that Denmark, a country providing universal health care coverage, ranks lower than the USA in responsiveness even though in the USA seven times as many people as live in Denmark have inadequate health insurance.[x] Navarro also expresses surprise that the USA ranks highest overall in terms of responsiveness and Colombia ranks highest on this measure in Latin America. Shaw has questioned whether it is plausible that Colombia has a fairer system of financing than Canada.[xi]

In response, Murray et al. [xii] have argued that Williams' interpretation of face validity is flawed as it fails to look at the concept of efficiency. Thus, they argue, it is entirely plausible that some of the rankings are explicable on the basis of differences in outcomes compared to resources invested.

In a more light-hearted review, albeit one underpinned by serious analysis, Appleby has noted how the rankings of overall performance generated in WHR 2000 are similar to those developed by FIFA to rank performance of national football teams.[xiii] Appleby argues that this may be because many of the determinants of what is being measured in the WHR 2000 are actually similar to those that determine the performance of national football teams, such as general economic development. If so, he argues, this calls into question whether the factors being captured by WHR 2000 really are measuring health system performance.

Questions:

a) *How important is face validity, as illustrated by the above examples, in ensuring the acceptability of a future assessment of health system performance?*
b) *While the examples cited are relatively uncontentious, many others are likely to be more controversial, potentially offering scope for endless arguments about rankings. If an element of face validity is to be incorporated, how can it most easily be done?*

Coherence of performance measures

The WHR 2000 sets out to measure the performance of the health system, which it defines broadly as including all individuals, groups, organisations and associated resources whose primary intent is to improve health. Thus, it includes, for example, policies to make roads safer and policies to change the national diet, as well as personal health services. It excludes actions whose indirect effect is to improve health, such as education. However, it has been argued that the measures of performance, based on available data, do not always relate specifically to this definition.

Thus, Navarro has argued that it is inappropriate to attribute differences in overall mortality (or DALE) solely to the health care system, quoting several passages from the report to make his point.[iv] For example, the WHR 2000 states: "If Sweden enjoys better health than Uganda – life expectancy is almost exactly twice as long – it is in large part because it spends exactly 35 times as much per capita in its health systems". He argues that no evidence is presented to support these claims, while there is considerable evidence that levels of mortality are determined by political interventions, and wealth and income distribution.

In response, Murray and Frenk concede that the evidence linking health interventions and outcomes is limited. The approach they took was to estimate what a country could be expected to achieve on the basis of its expenditure on health systems and education levels.[viii] This raises the question of how much a Ministry of Health can be held accountable for wider determinants of health. They address this using the example of tobacco consumption, a major determinant of premature mortality, arguing that Health Ministries must ultimately be held accountable for rates of smoking in their country, invoking the concept stewardship.

In turn, Navarro accepts that health care does affect health outcomes but restates his view that the primary determinants of population health lie outside the health care system and argues that his main point, that overall levels of health outcomes should not be used to assess performance of health systems, remains unanswered.[ix]

A second concern, voiced by, among others, the Vietnamese Ministry of Health, is the need to take account of the time lag between an intervention and a measurable impact on health.[xiv] The importance of this issue is growing with increasing understanding of the way in which factors acting in early life affect adult health. Thus, they argue, measures based on mortality now (notwithstanding the intrinsic delay in obtaining data less than a few years old even where such data are available) will at best reflect features of the health care system up to several decades previously.

The problem of linking performance measures to the definition of the health system that is used in WHR 2000 has also been noted by McKee.[xv] He attributes a greater role to health care in reducing mortality (at least in industrialised nations) than does Navarro but also notes that this role is limited to mortality from certain causes, and even there it is often difficult to differentiate variations in incidence and treatment outcome. He thus argues that, at least, a measure of performance should be limited to those causes of death where a direct link with the health system (however defined) can be made. While supporting the adoption of a broad, inclusive definition of the health system, embracing road safety and nutrition policy, he argues that it is difficult to relate fairness of financing to this goal and, while the measure of responsiveness claims to do so, it is likely that many informants will focus on the personal health care system.

Ugá et al. draw attention to the problem of linking performance to health systems in countries with multiple systems.[ii] Thus, they note, while Brazil scores poorly on responsiveness, surveys indicate a high level of satisfaction among users of the public system.

Questions:

a) Are the measures of performance used in WHR 2000 congruent with the definition of the health system that is used? Specifically, is the definition of health achievement, involving activities outside the personal health care system, congruent with the measures of fairness of financing and responsiveness, which appear to imply a narrower definition?

b) How helpful is the concept of stewardship? What, in practice, are its limits? To what extent can Ministries of Health be held accountable for overall levels of population health?

c) Is it possible to relate health outcomes more directly to the health system, taking account of the complex causal pathways leading to disability and premature death and the time lags between interventions and outcomes?

Data availability

Several commentators have drawn attention to the problems that arise because of lack of data from many countries. In the tables of measures that contribute to the overall index of attainment, but not the table of the composite index, the WHR 2000 identifies those values that are estimates by putting them in italics. Thus, Almeida et al note that the WHR 2000 was unable to obtain data from which to calculate the index of health inequality for 133 of the 191 countries (70%).[vi] The corresponding figure for the two measures of responsiveness was 161 countries (84% of the total). They were unable to determine the figures for levels of health. They also argue that the methods and underlying assumptions used to input missing values are inadequately specified and have not been subjected to peer review.

Williams makes similar points and reviews critically the ways in which missing data were estimated.[x] He also refers to his previous critical comments of the methods used in the Global Burden of Disease (GBD) study, on which levels of health are based. He concludes that "the composite index for attainment … is based on very little actual data, which is often heavily manipulated to make it usable, and then subjected to a great deal of rather adventurous modelling to fill out the rather large canvas of world health which the report purports to cover. A more virtuousic display of skating on thin ice you are unlikely ever to witness."

Muray et al. have responded by arguing that the use of estimates is common in international economic comparisons, in particular in calculation of GDP, GNP, purchasing power parities, and population numbers.[xii] They argue that taking a purist approach would mean rejecting anything based on such figures and that they, unlike others, at least use italics to indicate where estimates have been used.

Reflecting concerns about data, the WHR 2000 has made extensive use of estimates of uncertainty. This has been done using standard statistical methods based on an extended binomial model. Some commentators have, however, questioned where the boundary lies "between uncertainty and ignorance" (McMichael T, personal communication).

Questions:

a) *The framework adopted to assess performance is extremely demanding of data, much of which is unavailable for many countries. What is the appropriate response? Should missing data be omitted? What amount of basic data are necessary to permit imputation of missing values? How should questions of quality of existing data be handled?*
b) *Is the use of statistical sampling methods an appropriate way to assess uncertainty in the absence of primary data?*

Technical considerations in measuring overall performance

Health levels and distribution

Ugá et al. criticise the claims made for the measure of level of health.[ii] Disability-Adjusted Life Expectancy (DALE) is advocated by the authors of WHR 2000 as a superior measure to the simpler life expectancy at birth because it incorporates measures of disability. However, as data on levels of disability are available in very few countries, life expectancy elsewhere has been reduced by a factor that is the same (within each age band) for each country within groups that have similar levels of life expectancy. In countries where disability surveys are available, there was frequently a considerable mismatch between the values obtained from the surveys and those estimated using the GBD methodology. This was attributed by the authors to national differences in norms and expectations and thus the GBD data was used to rescale the survey data.[xvi] Consequently, the correlation between DALE and life expectancy at birth is 0.996 and Ugá et al. argue that the use of DALE adds nothing to the simpler, and arguably more transparent, measure of life expectancy at birth.

In response, WHO argues that the best available data for each country were used to estimate DALE, creating a non-trivial difference in DALE in countries with the same level of life expectancy at birth.[iii]

Almeida et al.[vi] also draw attention to previous concerns[xvii] about the health state valuations and disability weights employed in the generation of DALE.

Question:

Given the scarcity of data on disability, the many assumptions required, the concerns about the methods used to construct DALE, and the close association between DALE and life expectancy at birth, is there any benefit to be gained by using DALE rather than life expectancy at birth?

Other concerns have addressed the way in which inequality in health is measured used in the WHR 2000. In brief, the WHR 2000 uses a measure derived from the magnitude of differences in health between all individuals in the population, while much traditional work on inequalities considers differences between groups that have been pre-defined, based on known relationships between particular socio-demographic factors and health. These may include social class (variously defined), income, education, or ethnicity. Murray et al.[xviii] have previously argued that the former method has the following advantages: individuals with the worst health can be identified without pre-specifying social groups. Second, and arguably of most relevance here, comparisons of populations over time or of different populations are straightforward, and not susceptible to changes in group composition. Third, by separating the definition and measurement of inequality from pre-existing hypotheses then inequality itself becomes an object of scientific enquiry.

Braveman et al. have challenged this approach,[xix] although some of the points they make reflect the proposal, advocated initially by Murray et al.,[xviii] to assess differences in health among small geographical areas. They argue that the geographical groups, like social groups, involve an a priori specification of categorising variables. Observation of health differences among groups defined by, for example, education or social class does, however, prompt important questions about the pathways involved. In contrast, the mere observation of variations between individuals does not suggest even a general direction in which to look for explanations. They also argue that study of social inequalities does not prejudge mechanisms of causality any more than does the study of inequalities between geographical areas. Furthermore, they argue that the study of social inequalities not only does not prejudge causation but it ensures that such factors are explicitly considered.

They accept that differences in occupations and status over time and between countries make comparison difficult but argue that this should stimulate greater efforts to achieve clarity rather than simply "abandoning social description in favour of a technical solution that discards the key questions".

They conclude that the key issue is not whether a previously defined social categorisation should be used but whether the categorisation that is used is helpful in explaining inequalities in health. They further argue that the risk of obscuring important information is greater when an appropriate categorisation is not used than when it is.

In response, Murray et al. accept that geographical and social groupings both involve a priori categorisation of populations.[xx] They dispute the suggestion that studies of intra-individual differences will displace attention from social inequalities, arguing instead that it will help place inequalities per se on the international agenda. They reiterate that a focus on certain social determinants may lead to others being ignored and they restate the commitment of WHO to work on health inequalities.

The concerns of Braveman et al. have also been voiced by Almeida et al.,[vi] who also express concern that a technical paper cited in the WHR 2000 argues that an advantage of intra-individual variation in health is that it draws attention to worse health among the advantaged, if it exists, arguing that this is an issue equally important to health ministries as poor health among the disadvantaged. They argue that this view has troubling ethical and social policy implications.
Wolfson and Rowe make similar points.[xxi] They describe both intra-individual and geographic variations as univariate and social group classifications, in which mortality is related to a second, predetermined variable, as bivariate. They first address the use of geographical areas (a method proposed for, but not used in WHR 2000 following the decision to use child mortality) and show, by means of a simulation model, how the proposed method using data based on observed mortality rates cannot differentiate areas containing hypothetical populations with very different patterns of mortality. They take two populations, one containing individuals, all of whom have a risk of death of 0.5 in the period in question, and one where one half the population have a probability of death of 0.25 and the other have a probability of 0.75. The second population is clearly more unequal. However, in a series of simulations using data on 50,000 people aggregated into between 2000 and 25 groups, no consistent difference in standard deviations between the two populations is observable. This, they note, is because of the stochastic nature of death, leading to noise that obscures the true underlying relationship. They then examine the system used in the WHR 2000, which they note, uses data from the Demographic and Health Surveys (DHS) and thus has mothers as the main sampling unit. Here, they argue, it is plausible that certain conditions, such as environmental pollution, could cause both a failure of pregnancies to proceed to term and a high childhood mortality. Conversely, good circumstances could lead to larger families. In the extreme former case, with no surviving pregnancies, adversity will be missed. Thus, they argue, the univariate approach is intrinsically unable to detect differences in the distribution of mortality within populations.

Their second concern relates to the goal of capturing the risks to individuals over the course of their lifetime. They propose a series of possible approaches although all would require longitudinal data, which will only be available in a relatively small number of high-income countries. Their preferred model is a multi-state life table method in which all individuals in a population are tagged with a "frailty" or "resilience" index drawn from the estimated distribution of particular socio-economic variables at each point in life. By using a reference distribution of socio-economic factors they argue that it would be possible to compare inequalities between countries.

Ugá et al. also use simulation modelling to explore the consequences of contemporaneous changes in both effectiveness of health systems and distribution between social groups.[ii] Using the formula for distribution of health employed in WHR they simulate a population containing five social groups, each experiencing a higher life expectancy than the one below it. They assume that a change in health system increases life expectancy by one year for the three most affluent groups but by two years for the poorest two groups. At the same time, other social policies lead to an increase of 5% in the percentage of the poorest group. Although the health system change has reduced inequality, the net effect on the index of inequality is to increase it slightly. In a second simulation, comparing two populations, both with a high proportion (60%) in the poorest of five social groups, they show how, while a simple inspection of the variation in life expectancy demonstrates that one population is more unequal than another (range 55-68 years vs 58-66 years), the value of the index of inequality obtained indicates the opposite result.

The WHO has, however, challenged the inferences drawn from these examples.[iii] They show that, in the first example, the net effect of the changes is to increase the fraction of the population at

the extremes of the distribution, thus justifying the increase in the index of inequality. Similarly, the changes in the second example lead to a greater dispersion of life expectancies. Thus, it would appear that both groups correctly describe the changes that occur but then attribute different meanings to them.

In view of the dominance of social group analysis in the international literature on health inequalities and the inevitable tendency to associate the two approaches, Houwelling et al. have compared the rankings in the WHR 2000 with measures of social inequalities in countries where such data exist.[xxii] The measures studied were mortality among adult men from 15 high-income countries and among children in 43 low-income countries. In high-income countries a significant relationship was found, using both education and manual/ non-manual classifications, but this was based entirely on the existence of two outliers, Hungary and Estonia. When these two values were removed the relationships disappeared. Among children in low-income countries, the WHR 2000 measure correlated positively with the absolute rich-poor gap in mortality but negatively with the rich-poor mortality ratio. This did, however, appear to be due to the close correlation between the WHR 2000 measure and the absolute level of child mortality, which seems inevitable given the way that the figure for the distribution of child mortality is estimated.

Questions:

a) What are the relative merits of measures of inequality in health based on intra-individual and social group differences?

b) Are the methods proposed by Wolfson et al. useful for van exercise such as WHR 2000, given the scarcity of data in many countries and the contextual nature of socio-economic determinants of health?

c) If social group differences are to be addressed explicitly, how might this be done, taking account of coherence, context and comparability?

Responsiveness levels and distribution

The measures of responsiveness have attracted particularly widespread criticism because of the reliance on data from 1791 key informants in only 35 countries. Williams makes the following points.[x] The countries represented were atypical, excluding North America and Western Europe. Of the 42 questions asked of the respondents, only seven were used in the final measure, raising questions of why these were selected. Values from some countries were rejected because there was "no rational explanation" for the values obtained, with "systematic variations, not attributed to actual responsiveness" but, as Williams notes, if *actual* responsiveness was known, what was the purpose of undertaking the survey? Ugá et al. have also noted that, while it was intended that 50 key informants from each country would be recruited, in some countries (such as Brazil) the actual number was somewhat less.[ii]

Finally, as it was found that men, those in government employment, and those in countries with "less political and social freedom" tended to allocate higher scores to their countries, their scores were adjusted to produce the values that would be expected from a population that was 50% male and 50% female, who live in a totally free society, and who do not work for government. In its response to the WHR 2000, the Ministry of Health of Vietnam has drawn attention to the lack of criteria for political freedom or information on which countries are deemed to be politically free.[xiv]

Almeida et al have questioned the representativeness of informants, noting that they were exclusively professionals and half were WHO staff.[vi] They were unable to locate any discussion of the potential biases introduced.

Williams continues by questioning how the data from the 35 countries from which data exist have been extrapolated to the remaining countries of the world.[x] This involved constructing a regression model to explain the values obtained in the survey by means of more widely accessible data. He notes that, in doing so, data from India and China were omitted as they were not considered sufficiently representative of their populations, as do a further three because of particular national circumstances. Using these 30 countries, equations for each of the seven dimensions of responsiveness were generated using some of the following variables: % of population over 65; average years of schooling; geographic access rate; health expenditure per capita; GDP per capita; and percentage of private health expenditure. R-squared values were in the range 0.2-0.6.

The distribution of responsiveness was obtained by asking the key informants to identify groups that were often discriminated against in terms of responsiveness in their country. Four groups were identified most often: the poor; women; the old; and certain ethnic groups (whose identity varied according to the country concerned). An overall measure of distribution was generated by applying the frequency with which each group was mentioned by their frequency in the population of the country concerned. Once again missing values were generated from a regression equation based on the data points that were available, in this case using the percentage of the population below the poverty line and the proportion of the population living within one hour's travel of a health facility as explanatory variables.

Blendon et al. compared the WHR 2000 measures of responsiveness with data from surveys in seventeen industrialised countries.[xxiii] They found little association between the two measures, with the two countries rated most highly by their citizens appearing at the bottom of the WHO ranking.

In response, Murray et al.[xxiv] argue that Blendon et al. erred on two issues. First, the WHO informants were considering the broad definition of the health system used in the WHR 2000 while other surveys focused on personal health services. Second, the WHR 2000 addresses the efficiency with which resources are used, while the surveys did not. They continue by exploring reasons for rejecting simple measures of satisfaction as a means of assessing responsiveness. They note that expectations often differ, citing the lack of a relationship in several industrialised countries between the time that people spent in their last medical appointment (which they imply is a measure of quality) and satisfaction with the health system. In addition, in several countries, the poor report higher satisfaction with health services even though other evidence suggests they receive a less good service. They further argue that satisfaction fails to capture the full range of issues covered by the eight domains within the WHO responsiveness instrument (dignity, autonomy, confidentiality, information, prompt attention, access to social support networks, quality of basic amenities, and choice).

Murray et al. have also responded to the criticism that the ratings of responsiveness were derived from key informants by noting that, while time did not permit it for the WHR 2000, the instrument has subsequently been used in household surveys in over 70 countries and the results will be used in future rounds. They also describe the steps being taken to calibrate the instrument in view of the differences in norms and expectations that exist as, they argue, measures of satisfaction are heavily influenced by expectations. Thus, they argue, such surveys are

inappropriate. Instead they describe their work that asks respondents to rate a series of vignettes to explore the meaning attributed to different response categories.

In response, Blendon et al. repeat their argument that publication of the rankings should have awaited collection of the data now being gathered, that certain rankings lack face validity (quoting Italy, which scores highly although surveys reveal a high proportion of the population dissatisfied with the system), and that the differences in satisfaction observed in existing surveys are too great to be explained by differences in expectations.[xxv]

Question:

There appears to be a consensus that the method used in WHR 2000 was problematic. Is the method advocated for future rounds, using the eight domains already identified, and based on household surveys, with calibration using vignettes, an appropriate way forward?

Fairness of financing

Navarro and Almeida et al. have both criticised the definition of fairness of financing employed.[iv,vi] A fair financing system is defined as one in which different income groups spend the same percentage of their income on health care (after expenditure on food). This is criticised on two counts. One is that it takes no account of the frequently greater health needs of the poor, or the absolute cost involved in obtaining necessary care. Thus, a situation would be fair where a family with a disposable income of 1,000 received 100 of health care and a family with a disposable income of 100,000 received 10,000 worth of care. In each case, health expenditure accounts for 10% of household income but the care that could be provided would be manifestly different. Second, it takes no account of redistribution of resources. Thus, it is argued by Navarro and Almeida et al., as the poor will rarely be able to generate sufficient of their own resources to meet their legitimate health needs, a system is fair only when it permits transfers of resources for health care from the wealthy to the poor.

Question:

Should fairness of financing require an element of redistribution and, if so, how might this be assessed?

Estimating performance

The WHR 2000 measures overall performance against what is theoretically achievable. This is done by estimating what might be achieved in the absence of a health system, what has been achieved, and what could be achieved given the resources available. The WHR 2000 argues that, in the absence of a health system, distribution of responsiveness and fair financing have no meaning so they are given maximum weighting. This means, in effect, that the minimum value of the index is 37.5, equivalent to the weighting of these two measures. Williams questions why this is done but suggests that it is a relatively unimportant issue given the arbitrariness intrinsic to the overall exercise.[x] He also questions the system for deriving the minimum level of DALE for a given country. This is estimated as what would be predicted when the current level of educational attainment is applied to an equation of the relationship between literacy rates and levels of health in 1908. This, as he notes, implies that, other than health care, education is the principle determinant of DALE. This concern also applies to the derivation of the frontier for achievement of each country, which is also derived from education levels.

Question:

Is the method for assessing minimum and maximum achievable levels of performance valid? Are there any alternatives?

Use of a composite index

Almeida et al. argue that a composite index combining such disparate measures as health outcomes and fairness of financing is both difficult to understand and unhelpful as a guide to policy makers.[vi] They contrast them with other, more easily understood, indicators such as the UNDP's Human Development Index.

They have argued that the overall rankings of performance are highly sensitive to the weightings selected. The weightings used in the WHR 2000 rankings had been rounded to the nearest 12.5% from the original values obtained in the survey of key informants. When Almeida et al. recalculated the overall performance measures using the unrounded values, half of the countries moved by up to five ranks up or down. In response, the WHO has argued that few countries changed by more than one or two ranks and this was within the degree of uncertainty specified in the WHR 2000.[iii]

Question:

Is a composite index of performance meaningful or useful?

Other considerations

Use of evidence

Several commentators have argued that the WHR 2000 relies on an unduly narrow evidence base. Navarro's criticism of the ideological focus of those cited has been mentioned earlier.[iv] Almeida et al. draw attention to the fact that, in the statistical annex to the WHR 2000, only two of 32 citations are not by the authors of the report and 26 are in non-peer reviewed internal WHO documents. Williams makes a similar point about peer review, arguing that there is need for WHO to create a much more open intellectual environment for its staff, "with critics welcomed in and argued with (and even learned from)".[x]

Question:

Has the system of peer review and involvement of the wider research community undertaken so far been adequate? If not, how might it be improved in the future?

Opportunity cost

Several commentators have questioned whether the resources consumed by the production of the WHR 2000 might not have been used more effectively in other ways.

Williams asked whether "the analytical talents employed on this task [might better be diverted] into promoting an evaluative culture in countries that lack one so that they could find better solutions to their own problems?" [x] Braveman et al. have argued that the approach to health inequalities adopted in the WHR 2000 risks displacing efforts by WHO to address social inequalities in health.[xix]

Question:

Are the resources required for this exercise justified in relation to the value of the final product to Member States and in the light of competing goals such as strengthening the evidence base for policy and in-country capacity to use it?

[i] World Health Organization. World Health Report 2000. Health Systems: Improving Performance. Geneva: WHO, 2000.

[ii] Ugà A, Szwacwald CL, Almeida C et al. Views on the WHO 2000 Report. Rio de Janeiro: FIOCRUZ, undated.

[iii] WHO. 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, 17th and 19th January 2001. Geneva: WHO, 2001.

[iv] Navarro V. Assessment of the World Health Report 2000. Lancet 2000; 356: 1598-601.

[v] Navarro V. The new conventional wisdom: an evaluation of the WHO report Health Systems: Improving Performance. Int J Health Services 2001; 31: 23-33.

[vi] Almeida C, Braveman P, Gold MR, et al. Methodological concerns and recommendations on policy consequences of the World Health report 2000. Lancet 2001; 357: 1692-7.

[vii] Helms R. Sick list: health care à la karl Marx. Wall Street Journal. Brussels: June 29, 2000.

[viii] Murray C, Frenk J. World Health Report 2000: a step towards evidence-based health policy. Lancet 2001; 357: 1698-700.

[ix] Navarro V. World Health Report 2000: responses to Murray and Frenk. Lancet 2001; 357: 1701-2.

[x] Williams A. Science or marketing at WHO? A commentary on "World Health 2000". York: centre for Health Economics, 2001

[xi] The Lancet. Why rank countries by health performance? Lancet 2001; 357: ##

[xii] Murray CJL, Frenk J, Evans D, Kawabata K, Lopez A, Adams O. Science or marketing at WHO? A response to Williams. Geneva: WHO, undated.

[xiii] Appleby J. Health Service Goals: Life, death and football J Health Serv Res Pol (in press)

[xiv] Ministry of Health. Comments and suggestions of Vietnam Ministry of Health/ Health Policy Unit as regards the World Health Report 2000. Hanoi: Ministry of Health, undated.

[xv] McKee M. Measuring the efficiency of health systems. BMJ 2001; 323: 295-6.

[xvi] Mathers CD, Sadana R, Salomon JA, Murray CJL, Lopez AD. Healthy life expectancy in 191 countries, 1999. Lancet 2001; 357: 1685-91.

[xvii] Anand S, Hanson K. Disability-adjusted life years: a critical review. J Health Econ 1997; 16: 685-702.

[xviii] Murray CJL, Gakidou EE, Frenk J. Health inequalities and social group differences: what should we measure? Bull WHO 1999; 77: 537-452.

[xix] Braveman P, Kreiger N, Lynch J. Health inequalities and social inequalities in health. Bull WHO 2000; 78: 232-233.

[xx] Murray CJL, Gakidou EE, Frenk J. Response to P. Braveman et al. Bull WHO 2000; 78: 234.

[xxi] Wolfson M, Rowe G. On measuring inequalities in health. Bull WHO 2001; 79: 553-560.

[xxii] Houwelling TAJ, Kunst AE, Machenbach JP. World Health Report 2000: inequality index and socio-economic inequalities in mortality. Lancet 2001; 357: 1671-2.

[xxiii] Blendon RJ, Kim M, Benson JM. The public versus the World Health Organization on health systems performance. Health Affairs 2001; 20:

[xxiv] Murray CJL, Kawabata K, Valentine N. People's experience versus people's expectations. Health Affairs 2001; 20 (May/ June): 21-23.

[xxv] Blendon RJ, Kim M, Benson JM. Authors respond to WHO critics. Health Aff 2001; 20: 253-4.