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.

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.